

FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR DE CIÊNCIAS DO TRABALHO E DA EMPRESA

Departamento de Matemática da FCUL

Departamento de Finanças do ISCTE



**UNIVERSIDADE
DE LISBOA**

ISCTE  **Business School**
Instituto Universitário de Lisboa

**COINTEGRAÇÃO, MODELOS VAR E BVAR:
ESTUDO COMPARATIVO ENTRE A ABORDAGEM CLÁSSICA E BAYESIANA
NO CONTEXTO DOS MERCADOS FINANCEIROS EUROPEUS**

Filipe Roberto de Jesus Ramos

MESTRADO EM MATEMÁTICA FINANCEIRA

2011

FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR DE CIÊNCIAS DO TRABALHO E DA EMPRESA

Departamento de Matemática da FCUL

Departamento de Finanças do ISCTE



**UNIVERSIDADE
DE LISBOA**

ISCTE  **Business School**
Instituto Universitário de Lisboa

**COINTEGRAÇÃO, MODELOS VAR E BVAR:
ESTUDO COMPARATIVO ENTRE A ABORDAGEM CLÁSSICA E BAYESIANA
NO CONTEXTO DOS MERCADOS FINANCEIROS EUROPEUS**

Filipe Roberto de Jesus Ramos

MESTRADO EM MATEMÁTICA FINANCEIRA

Dissertação Orientada pela

Professora Doutora Diana Aldea Mendes

Departamento de Métodos Quantitativos, ISCTE – IUL

2011

RESUMO

Tendo em vista o desenvolvimento de um estudo compreendendo modelação econométrica, o objectivo central deste trabalho consiste na análise comparativa de resultados obtidos na estimação de um sistema cointegrado de vectores auto-regressivos e da respectiva ordem de cointegração, sob o ponto de vista dos pressupostos das abordagens clássica (VAR) e bayesiana (BVAR). Para o efeito, considerou-se uma amostra de variáveis provenientes de índices bolsistas de seis países da zona Euro.

Antecedendo à implementação prática dos modelos econométricos, é apresentado, um referencial teórico que visa fornecer, de forma gradual, informações consideradas indispensáveis à compreensão dos modelos econométricos em causa. Assim, além dos aspectos inerentes à abordagem clássica, são analisados os métodos de simulação de Monte Carlo via Cadeias de Markov (MCMC), apresentando-se o algoritmo de *Metropolis-Hastings* e o método de amostragem de *Gibbs*, cujos fundamentos foram discutidos com base na teoria de Cadeias de Markov.

Estes métodos, baseados em simulação estocástica iterativa, constituem a base da inferência bayesiana, a qual se fundamenta no conhecimento da distribuição *a posteriori* dos parâmetros e na possibilidade de construção de intervalos de confiança exactos para as estimativas dos parâmetros considerados. Precisamente, sendo o problema na implementação desta abordagem a integração numérica (no espaço do parâmetro) da fdp *a posteriori*, a opção para resumir a informação descrita na distribuição *a posteriori* foram os métodos MCMC.

As vantagens reconhecidas na implementação das técnicas bayesianas têm sido claramente identificadas em estudos ulteriores, daí o recurso bastante comum às mesmas, pois problemas anteriormente intratáveis foram solucionados e modelos mais flexíveis foram introduzidos com elevado sucesso.

Palavras-chave: Inferência Bayesiana; Cadeias de Markov; Simulação MCMC; Cointegração; Modelos VAR e BVAR.

Classificação JEL: C11, C32, C58, G15

ABSTRACT

Given the development of an econometric modeling study, the main objective of this work consists in the comparative analysis of results obtained in the estimation of a cointegrated vector autoregressive model and its cointegration rank, from the point of view of the assumptions of the classical (VAR) and Bayesian (BVAR) approaches. For this purpose a sample of variables of stock indices from six countries of the Euro area, were taken into account.

Preceding to the practical implementation of the econometric models, we present a theoretical framework that aims to provide, in a gradual manner, information deemed necessary for the understanding of the econometric models in question. Thus, besides the inherent aspects of the classical approach, the Markov Chains Monte Carlo (MCMC) simulation methods are analyzed, by presenting in particular the Metropolis-Hastings algorithm and Gibbs sampling method, whose foundations were discussed based on Markov Chains theory.

These methods, based on iterative stochastic simulation form the basis of bayesian inference, relying on the knowledge of the posterior distribution of parameters and on the possibility of constructing exact confidence intervals to estimate the considered parameters. Precisely, since the main tool of the implementation of this approach consists the numerical integration (in the parameter space) of the posterior fdp, the option to summarize the information described in the posteriori distribution are based on the MCMC methods.

The advantages in the implementation of Bayesian techniques have been clearly identified in further studies (observe the quite common appeal to them) hence previously intractable problems have been resolved and more flexible models were introduced with great success.

Keywords: Bayesian inference, Markov Chains, MCMC simulation, Cointegration, VAR and BVAR models.

JEL Classification: C11, C32, C58, G15

AGRADECIMENTOS

Pelo significado que a palavra agradecimento em si comporta, antecedendo a toda a apresentação deste trabalho, fará sentido dedicar algumas palavras a todos os que, de uma forma mais activa ou pelo simples apoio, ajudaram à sua concretização.

Assim, em momentos diferentes, um reconhecimento pela compreensão e apoio dos colegas de trabalho e dos elementos pertencentes a hierarquias superiores, quer da José Cardoso Pires (em especial dos elementos da CAP e das Equipas Pedagógicas dos CEF), quer da Fidelidade (particularmente, Cláudia B., Dr.^a Ana Conceição e Dr.^a Vera Barahona).

Com os colegas e companheiros de Mestrado, não só de turma (nomeadamente a Julieta, Isalinda, Inês, Margarida e Vítor), como outros que fui conhecendo (Bernardo, Davide, Lúcia, Madalena e Marta), sublinho os momentos de rica partilha de informação, opiniões e conhecimentos, alguns deles contribuindo para o enriquecer deste trabalho.

Agradeço, particularmente, a alguns amigos, Alexandra, Cristina, Fernanda, Karin, Mafalda, Mário e Ricardo, pelo contributo e sugestões na consecução deste trabalho.

Ao corpo docente do Mestrado (FCUL e ISCTE), uma palavra de gratidão pelos ensinamentos. Pelo mesmo motivo e pela disponibilidade, apoio e incentivo, um reconhecimento especial à Professora Adelaide Carreira, Professora Amélia Fonseca, Professora Gracinda Gomes, Professora Isabel Simão e Professora Purificação Coelho.

Mais do que merecido, pela infinita paciência, compreensão e apoio incondicional, um sincero agradecimento aos meus pais e irmão.

A última palavra, pelo contributo e dedicação determinantes, vai para alguém que me ajudou e incentivou de forma ímpar, sendo, seguramente uma referência pela elevada competência e profissionalismo, a minha orientadora, Professora Diana Mendes.

A todos, o meu sincero Obrigado.

ÍNDICE GERAL

RESUMO	I
ABSTRACT	III
AGRADECIMENTOS	V
ÍNDICE GERAL	VII
ÍNDICE DE FIGURAS	XI
ÍNDICE DE TABELAS	XIII
INTRODUÇÃO	1
1. REVISÃO DA LITERATURA E ENQUADRAMENTO TEÓRICO	5
1.1. ABORDAGEM BAYESIANA: ESTUDOS PRECEDENTES.....	5
1.2. INFERÊNCIA ESTATÍSTICA: CLÁSSICA E BAYESIANA.....	7
1.2.1. INFERÊNCIA CLÁSSICA.....	7
1.2.2. INFERÊNCIA BAYESIANA	10
1.2.3. TEOREMA DE BAYES	13
1.2.4. REFERÊNCIA À DISTRIBUIÇÃO <i>A PRIORI</i>	18
1.3. CADEIAS DE MARKOV EM TEMPO DISCRETO.....	19
1.3.1. NOÇÕES GERAIS EM CADEIAS DE MARKOV.....	22
1.3.2. DISTRIBUIÇÃO ESTACIONÁRIA: CONDIÇÃO DE EXISTÊNCIA, UNICIDADE E CONVERGÊNCIA	28
1.3.3. DAS CADEIAS DE MARKOV AOS MÉTODOS MCMC.....	31
1.4. MÉTODOS DE SIMULAÇÃO ESTOCÁSTICA.....	32
1.4.1. INTRODUÇÃO AOS MÉTODOS DE SIMULAÇÃO.....	32
1.4.2. MÉTODO DE MONTE CARLO (MMC)	33
1.4.2.1. MMC ORDINÁRIO	35
1.4.2.2. MMC COM AMOSTRAGEM DE IMPORTÂNCIA	38
1.4.3. MÉTODO DE MONTE CARLO VIA CADEIAS DE MARKOV (MCMC).....	40
1.4.3.1. ALGORITMO DE <i>METROPOLIS-HASTINGS</i>	42
1.4.3.2. MÉTODO DE AMOSTRAGEM DE <i>GIBBS</i>	45

2. SÉRIES TEMPORAIS: FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA	49
2.1. INTRODUÇÃO: MODELAÇÃO EM ECONOMETRIA.....	49
2.2. NOÇÕES GERAIS NO ESTUDO DE SÉRIES TEMPORAIS	52
2.2.1. ESTACIONARIEDADE.....	53
2.2.2. RAIZ UNITÁRIA	56
2.2.3. REGRESSÃO ESPÚRIA.....	59
2.2.4. COINTEGRAÇÃO.....	60
2.2.5. CRITÉRIOS DE COMPARAÇÃO / SELECÇÃO DE MODELOS	62
2.2.5.1. TESTE DA RAZÃO DE VEROSIMILHANÇA	63
2.2.5.2. CRITÉRIO DE INFORMAÇÃO DE AKAIKE	63
2.2.5.3. CRITÉRIO DE INFORMAÇÃO BAYESIANA.....	64
2.2.5.4. COMPARAÇÃO DOS CRITÉRIOS.....	65
2.3. MODELOS ECONÔMICOS: VAR E VECM	66
2.4. MODELOS ECONÔMICOS E COINTEGRAÇÃO: ABORDAGEM BAYESIANA.....	74
2.4.1. CONSIDERAÇÕES GERAIS NA ABORDAGEM BAYESIANA.....	74
2.4.2. DISTRIBUIÇÃO A <i>PRIORI</i>	78
2.4.3. DISTRIBUIÇÃO A <i>POSTERIORI</i>	81
3. ANÁLISE DESCRITIVA DAS SÉRIES TEMPORAIS.....	87
3.1. APRESENTAÇÃO DAS SÉRIES TEMPORAIS	87
3.2. ANÁLISE GRÁFICA E ESTATÍSTICAS DESCRITIVAS	88
3.3. ESTUDO DA ESTACIONARIEDADE	91
4. ESTIMAÇÃO DA ORDEM DE COINTEGRAÇÃO E MODELOS ECONÔMICOS ..	93
4.1. ESTIMAÇÃO/INFERÊNCIAS USANDO A METODOLOGIA CLÁSSICA.....	93
4.2. ESTIMAÇÃO/INFERÊNCIAS USANDO A METODOLOGIA BAYESIANA.....	95
4.3. ANÁLISE COMPARATIVA DOS RESULTADOS OBTIDOS	97
CONCLUSÃO	99
REFERÊNCIAS BIBLIOGRÁFICAS	103
ANEXO A.....	107
A.1. EXEMPLOS RELATIVOS A CADEIAS DE MARKOV EM TEMPO DISCRETO .	107
A.2. EXEMPLOS RELATIVOS AOS MCMC	111
ANEXO B	113

ANEXO C	115
C.1. GRÁFICOS E ESTATÍSTICAS DESCRITIVAS DAS SÉRIES (LOG-NÍVEL E DL)115	
C.2. CAUSALIDADE BIVARIADA DE GRANGER.....	118
ANEXO D.....	121
D.1. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2002 – 7:2010).....	121
D.2. MODELOS VAR/VECM (1:2002 – 7:2010).....	124
D.3. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2002 – 1:2008)	128
D.4. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2008 – 7:2010)	130
D.5. MODELOS VAR/VECM (1:2008 – 7:2010).....	134
ANEXO E.....	139
E.1. ANÁLISE BAYESIANA DA ORDEM DE COINTEGRAÇÃO (1:2002 – 7:2010)	139
E.2. ANÁLISE BAYESIANA DA ORDEM DE COINTEGRAÇÃO (1:2008 – 7:2010)	142
E.3. MODELOS BVAR/VECM (1:2008 – 7:2010)	146

ÍNDICE DE FIGURAS

FIGURA 1 – Esquema ilustrativo dos procedimentos na Inferência Clássica	12
FIGURA 2 – Esquema ilustrativo dos procedimentos na Inferência Bayesiana	12
FIGURA 3 – Representação gráfica do efeito da distribuição <i>a priori</i>	16
FIGURA 4 – Ciclo iterativo da construção de modelos de modelação (Box & Jenkins)	53
FIGURA 5 – Exemplos de séries estacionárias e não estacionárias	54
FIGURA 6 – Processos de transformação de uma série com vista à estacionariedade	55
FIGURA 7 – Exemplo de séries cointegradas e não cointegradas	60
FIGURA 8 – Esquema de implementação da Metodologia de Johansen	73
FIGURA 9 – Representação gráfica das séries logaritmizadas.....	88
FIGURA 10 – Representação gráfica das séries em DL.....	90
FIGURA 11 – Representação gráfica da relação de longo prazo (1:2002 – 7:2010).....	94
FIGURA 12 – Representação gráfica e estatísticas descritivas individualizada das séries logaritmizadas.....	116
FIGURA 13 – Representação gráfica e estatísticas descritivas individualizada das séries DL	117
FIGURA 14 – Representação gráfica relativa à estabilidade do modelo VECM (1:2002 – 7:2010).....	124
FIGURA 15 – Representação gráfica relativa à estabilidade do modelo VAR (1:2002 – 7:2010).....	126
FIGURA 16 – Representação gráfica da relação de longo prazo (1:2008 – 7:2010).....	134
FIGURA 17 – Representação gráfica relativa à estabilidade do modelo VECM (1:2008 – 7:2010).....	134
FIGURA 18 – Representação gráfica relativa à estabilidade do modelo VAR (1:2008 – 7:2010).....	136
FIGURA 19 – Representações gráficas relativas ao modelo BVECM.....	147
FIGURA 20 – Representações gráficas relativas ao modelo BVAR.....	147

ÍNDICE DE TABELAS

TABELA 1 – Esquema de computação do MMC Ordinário	38
TABELA 2 – Esquema de computação do MMC com amostragem de importância	40
TABELA 3 – Esquema de computação do algoritmo de <i>Metropolis-Hastings</i> (MCMC)	45
TABELA 4 – Esquema de computação do método de amostragem de <i>Gibbs</i> (MCMC)	46
TABELA 5 – Estatísticas descritivas das séries logaritmizadas	89
TABELA 6 – Matriz de Correlação entre as séries logaritmizadas	90
TABELA 7 – Resumo dos resultados relativos ao teste à existência de raiz unitária para as séries logaritmizadas	91
TABELA 8 – Resumo dos resultados relativos ao teste à existência de raiz unitária para as séries em DL.....	92
TABELA 9 – Resumo dos resultados relativos à estimação do número de <i>lags</i> óptimo e ordem de cointegração (abordagem clássica)	94
TABELA 10 – Resumo dos resultados relativos à estimação do número de <i>lags</i> óptimo e ordem de cointegração (abordagem bayesiana)	96
TABELA 11 – Resumo geral dos resultados relativos à estimação do número de <i>lags</i> óptimo e ordem de cointegração	98
TABELA 12 – Estatísticas descritivas das séries logaritmizadas (1:2008 – 7:2010).....	115
TABELA 13 – Matriz de Correlação entre as séries logaritmizadas (1:2008 – 7:2010)	115
TABELA 14 – Estatísticas relativas ao teste de causalidade bivariada de Granger.....	119
TABELA 15 – Estatísticas relativas à estimação do número de <i>lags</i> óptimo (1:2002 – 7:2010)	121
TABELA 16 – Estatísticas relativas ao sumário da ordem de cointegração para 1 <i>lag</i> (SC) (1:2002 – 7:2010).....	122
TABELA 17 – Estatísticas relativas ao teste de cointegração de Johansen – 1 <i>lag</i> – (<i>Intercept</i> e no <i>Trend</i>) (1:2002 – 7:2010)	123
TABELA 18 – Estatísticas relativas ao sumário da ordem de cointegração para 20 <i>lags</i> (AIC) (1:2002 – 7:2010).....	124

TABELA 19 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VECM (1:2002 – 7:2010)	126
TABELA 20 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VAR (1:2002 – 7:2010)	128
TABELA 21 – Estatísticas relativas à estimação do número de <i>lags</i> óptimo (1:2002 – 1:2008).....	129
TABELA 22 – Estatísticas relativas ao sumário da ordem de cointegração para 1 <i>lag</i> (SC) (1:2002 – 1:2008).....	129
TABELA 23 – Estatísticas relativas ao sumário da ordem de cointegração para 3 <i>lags</i> (AIC) (1:2002 – 1:2008).....	130
TABELA 24 – Estatísticas relativas à estimação do número de <i>lags</i> óptimo (1:2008 – 7:2010).....	131
TABELA 25 – Estatísticas relativas ao sumário da ordem de cointegração para 1 <i>lag</i> (SC) (1:2008 – 7:2010).....	131
TABELA 26 – Estatísticas relativas ao teste de cointegração de Johansen – 1 <i>lag</i> – (<i>Intercept</i> e no <i>Trend</i>) (1:2008 – 7:2010)	133
TABELA 27 – Estatísticas relativas ao sumário da ordem de cointegração para 8 <i>lags</i> (AIC) (1:2008 – 7:2010).....	133
TABELA 28 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VECM (1:2008 – 7:2010)	136
TABELA 29 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VAR (1:2008 – 7:2010)	138
TABELA 30 – Estimação bayesiana da ordem de cointegração 2 <i>lag</i> (SC) (1:2002 – 7:2010).....	140
TABELA 31 – Estimação bayesiana da ordem de cointegração 19 <i>lags</i> (AIC) (1:2002 – 7:2010).....	142
TABELA 32 – Estimação bayesiana da ordem de cointegração 1 <i>lag</i> (SC) (1:2008 – 7:2010).....	144
TABELA 33 – Estimação bayesiana da ordem de cointegração 8 <i>lags</i> (AIC) (1:2008 – 7:2010).....	146

INTRODUÇÃO

Partindo de informações reais, um dos propósitos fulcrais da Econometria, muitas vezes definida como “mensuração económica”, é estimar empiricamente relações económicas, cuja metodologia passa pela estimação de modelos – modelos econométricos – que não são mais do que uma representação simplificada do real, que procura representar, explicar e prever relações.

Porém, pensar que todo o processo de modelação é linear, é sinónimo de uma visão redutora do trabalho desenvolvido em Econometria, o qual contempla duas vertentes, uma teórica e outra aplicada. Se a vertente teórica envolve o desenvolvimento de modelos e métodos apropriados à referida mensuração das relações entre variáveis económicas comportando uma forte componente matemática e estatística; a vertente aplicada faz uso dessas mesmas teorias, implementando-as em determinados contextos e a dados reais, podendo estes ser séries temporais (*times series data*), dados teoricamente cruzados (*cross-sectional*) e dados em painel (*panel data*).

É precisamente nesta filosofia que assenta o trabalho que nos propomos desenvolver nas páginas subsequentes. Partindo de uma fundamentação teórica visando clarificar e dar consistência aos modelos econométricos em análise, far-se-ão inferências, prevendo relações entre variáveis, após a implementação prática dos modelos em causa a uma amostra de dados reais que, neste caso são constituídos por séries temporais de índices de preços de fecho criados pela Thomson Reuters® representativos do comportamento do mercado accionista de seis países da zona euro. Um dos aspectos que valoriza o trabalho desenvolvido, deve-se à análise comparativa de resultados obtidos na implementação de metodologias fundamentada na inferência clássica e na inferência bayesiana. Assim a referida valorização fundamenta-se no facto do estudo em causa não se encerrar com a perspectiva clássica de modelação econométrica e estudo de cointegração entre séries temporárias, mas sim pelo confronto de resultados e procura da existência, ou não, de vantagens quando o processo de modelação e estudo da cointegração é feito à luz dos pressupostos da inferência bayesiana.

Deste modo, para o estudo em causa, considerou-se a divisão do trabalho em cinco partes, sendo a última reservada para uma síntese das principais conclusões do nosso estudo.

Em linhas gerais, numa primeira parte (Capítulo 1) é feita uma revisão da literatura sobre o tema em estudo bem como um encadeamento teórico dos principais aspectos subjacentes ao estudo em causa. Com efeito, começar-se-á por fazer uma breve referência a alguns estudos que serviram de motivação inicial para o estudo do tema, onde é realçada a utilização dos pressupostos bayesianos no processo de inferência, destacando as vantagens comparativamente à abordagem clássica. Neste seguimento, no sentido de clarificar as duas abordagens será feita uma breve revisão, com base em Paulino *et al.* (2003), às principais ideias subjacentes à Inferência Clássica e Inferência Bayesiana, seguindo-se uma análise ao resultado que está na base da abordagem bayesiana – *Teorema de Bayes*, o qual incorpora dois conceitos fundamentais a explorar: distribuição *a priori* e distribuição *a posteriori*. Como a distribuição *a posteriori* será objecto de maior detalhe nas secções seguintes, esta secção termina com uma referência breve à distribuição *a priori*.

Implícita à abordagem bayesiana que nos propomos desenvolver, surge a necessidade de encontrar uma forma de obter a distribuição *a posteriori*, a qual se obtém por meio de simulações estocásticas, como são exemplo os Métodos de Monte Carlo. Segue-se então uma abordagem a estes métodos de simulação estocástica, iniciando-se com uma visão geral sobre o tema. Especificamente, sendo que a nossa opção foi os métodos de Monte Carlo baseados em Cadeias de Markov (métodos MCMC), cuja distribuição estacionária de interesse é a distribuição *a posteriori*, serão descritos dois procedimentos possíveis (entre outros): o algoritmo de *Metropolis-Hastings* e o método de amostragem de *Gibbs*. Porém, sendo um dos pré-requisitos à compreensão dos aspectos em causa, o conhecimento das ideias principais associadas às cadeias de Markov, a anteceder à apresentação dos métodos MCMC, é feita uma revisão (sucinta) a noções gerais em Cadeias de Markov e a alguns resultados teóricos, com especial destaque para os resultados que garantem a convergência da amostra gerada pelos métodos de MCMC para a distribuição de interesse.

Feito o enquadramento teórico, passando o nosso estudo pela análise de séries temporais, numa perspectiva de análise econométrica, no Capítulo 2 é feita uma fundamentação dos modelos econométricos implementados: VAR e VECM. Sendo esta uma parte crucial deste trabalho, a análise aos modelos em causa mereceu um duplo tratamento. Em alternativa a uma apresentação densa, optou-se por repartir a análise em duas secções onde são descritas, em linhas gerais, as ideias subjacentes aos modelos VAR e VECM, com particular ênfase para a análise da cointegração, segundo a perspectiva clássica (Secção 2.3) com base nos trabalhos desenvolvidos por Johansen (1988) e Johansen e Juselius (1990) e numa perspectiva bayesiana (Secção 2.4) fundamentada nos trabalhos desenvolvidos por Villani (2005) e Warne (2006).

A anteceder toda esta apresentação, impunha-se uma referência a noções fundamentais implícitas no estudo de séries temporais, como estacionaridade, fundamentada nos testes de raiz unitária, regressão espúria e cointegração (conceito chave do nosso trabalho), seguida de uma breve descrição dos critérios usados na selecção e comparação dos modelos a usar.

Os dois capítulos seguintes (Capítulos 3 e 4) respeitam à parte prática do nosso estudo, a qual se inicia com uma análise descritiva das séries temporais em estudo provenientes de Índices Bolsistas de seis países do Mercado Europeu. Já no Capítulo 4 é feita uma análise e confronto dos modelos VAR e VECM obtidos segundo uma abordagem clássica e bayesiana. Na implementação prática dos aspectos inerentes às duas abordagens feitas, foram utilizados dois suportes em termos de software. O Eviews foi a opção escolhida para o estudo segundo os pressupostos de inferência clássica, pelo que todas as representações gráficas e os outros outputs apresentados foram obtidos com a implementação das funcionalidades do mesmo. Para a abordagem bayesiana, toda a metodologia seguida teve por base a utilização de códigos de programação implementados em MatLab, desenvolvidos por Warne com base nos pressupostos explorados nos trabalhos supracitados.

Na quinta e última parte – Conclusão – são apresentadas e sintetizadas as principais conclusões deste estudo, como já havíamos referido, procurando estabelecer uma ligação (crítica) dos resultados apurados com a natureza dos dados em causa. São ainda tecidas algumas considerações relativamente a algumas limitações do estudo, bem como aspectos que poderão ser desenvolvidos em estudos ulteriores, partido do presente estudo.

No que respeita aos Anexos, no Anexo A, face aos dados usados no estudo, é feita uma breve referência à descrição do processo de indexação, como base na informação disponível no *site* da Thompson Reuters[®]. Enquanto nos Anexos B, C e D constam os *outputs* referentes à parte prática. Especificamente, no Anexo B constam mais alguns dados e gráficos referentes à análise descritiva das variáveis, que vêm fundamentar algumas das conclusões constantes no Capítulo 3. Já os anexos C e D respeitam a estimação do modelo e a ordem de cointegração, seguindo as duas abordagens em estudo, clássica (Anexo C) e bayesiana (Anexo D).

1. REVISÃO DA LITERATURA E ENQUADRAMENTO TEÓRICO

1.1. ABORDAGEM BAYESIANA: ESTUDOS PRECEDENTES

O paradigma bayesiano usa probabilidades para expressar a incerteza de tudo o que é desconhecido. Deste modo, as previsões bayesianas podem incorporar informações *a priori* resultantes de crenças sobre os parâmetros assumidos num modelo, ou resultantes da influência (através da função de verosimilhança) de dados históricos observados.

Compreender como tal informação é introduzida no processo de inferência, constituiu uma motivação notável para o estudo que nos propomos desenvolver, a qual acabaria por ganhar maior ênfase quando analisados alguns trabalhos que evidenciam uma notável aplicabilidade deste paradigma a estudos específicos em contexto real.

Não sendo possível uma referência proporcional, face aos inúmeros estudos existentes, enumeram-se alguns¹, procurando ilustrar a aplicabilidade dos modelos bayesianos (eventualmente a outras áreas de investigação que não a financeira) e a forma como a abordagem bayesiana lida com diversos desafios empíricos. Estes desafios incluem, por exemplo, a modelação das variáveis financeiras, tal como a volatilidade, onde existe a necessidade de se lidar com múltiplas fontes de não-linearidade.

Hoogerheide e van Dijk, em "*Bayesian Forecasting of Value at Risk and Expected Shortfall Using Adaptive Importance Sampling*", centraram-se na previsão da distribuição dos retornos dos activos, cujo principal objectivo incidiu no estudo da cauda inferior da distribuição (associado às perdas) e na estimação adequada de previsões bayesianas do *value at risk* (VaR) e do *expected shortfall* (ES).

Com particular interesse pela parte computacional, Zellner e Ando no trabalho "*Bayesian and Non-Bayesian Analysis of the Seemingly Unrelated Regression Model with Student-t Errors and its Application for Forecasting*", desenvolveram um método computacionalmente eficiente para a aplicação da inferência bayesiana a modelos de regressão. Os autores sugerem um novo algoritmo que combina Monte Carlo directo com amostras importantes, com o intuito de calcular diversas quantidades de interesse. Os

¹ Os artigos que não estão devidamente identificados com a respectiva data de publicação, estão citados no trabalho de Lahiri e Martin (2010).

autores apresentam ilustrações numéricas, incluindo a produção de previsões de vendas para dez sectores industriais no Japão.

Mostrado a aplicabilidade da metodologias noutras áreas, podemos referir o trabalho desenvolvido por Yelland, *“Bayesian Forecasting of Parts Demand”*, que estuda uma solução bayesiana para o problema de prever a procura de peças de computador com pouco, ou mesmo nenhum histórico de procura, estando aqui em foque uma das limitações conhecidas da inferência clássica (estimação com amostras muito pequenas).

Já Costa (2004) aborda o estudo de técnicas bayesianas de inferência estatística, apresentando a aplicabilidade de três algoritmos de integração numérica estocástica, via simulação Monte Carlo em Cadeias de Markov (os algoritmos de Metropolis-Hastings e Reversible Jump e o método de amostragem de Gibbs), na selecção da ordem de um modelo auto-regressivo. A abordagem proposta é ilustrada com dados sintéticos e com um sinal de áudio.

Continuando na linha de estudo aplicados a vectores auto-regressivos (VAR), caso com particular interesse para o nosso trabalho, Koop e Strachan, em *“Bayesian Forecasting using Stochastic Search Variable Selection in a VAR subject to Breaks”*, apresentam um modelo VAR que é aumentado para providenciar um número desconhecido de pausas, sendo que cada pausa pode ocorrer com probabilidade constante e em cada período de tempo. As previsões são produzidas como uma distribuição mista, onde os diversos pesos são as probabilidades atribuídas ao acontecimento "fazer pausa" e ao acontecimento complementar, ocorridos durante o horizonte da previsão.

Jaramillo (2008), aplicando as metodologias bayesianas, estuda os mecanismos de transmissão da política monetária, efectuando projecções para as principais variáveis macroeconómicas para a economia chilena. Os resultados obtidos pela metodologia bayesiana são confrontados com os resultantes da implementação da metodologia clássica, discutindo-se algumas implicações para a política monetária.

Finalmente, Félix e Nunes (2002), no trabalho desenvolvido apresentava-se na altura, segundo os autores, uma visão pioneira, na medida em que estudos antecedentes sobre modelos VAR bayesianos desenvolvidos para a zona euro não incorporavam a modelação explícita de relações de longo prazo, facto que não acontece no trabalho em questão. Assim, o objectivo imediato desta aplicação para a zona euro é a previsão de um conjunto de variáveis que habitualmente desempenham um papel preponderante em exercício de previsões: o produto, o desemprego, a inflação, os salários, a taxa de juro e a taxa de câmbio.

Pelo contributo que outros estudos desenvolvidos pelos mesmos tiveram na consecução deste trabalho, alguns dos autores supramencionados, serão invocados em secções seguintes deste trabalho.

1.2. INFERÊNCIA ESTATÍSTICA: CLÁSSICA E BAYESIANA

É no seguimento do principal desígnio da estatística, “fazer inferências ou predições acerca de novos fenómenos ou de novas observações da mesma natureza” depois de feita uma análise exploratória, ainda que informal dos dados iniciais, (Robert *apud* Paulino *et al.*, 2003, p.1), que se enquadra um dos propósitos centrais deste trabalho.

Indo além dos procedimentos dito clássicos, no aprofundar de um estudo estatístico, encontramos várias correntes/escolas, cada uma com os seus princípios e procedimentos próprios: bayesianos (objectiva, subjectiva, ...), estruturalistas, verosimilhancistas (Paulino *et al.*, 2003). Como já foi referido, o estudo estatístico aqui desenvolvido terá como alicerce a inferência bayesiana, que considera a probabilidade como a medida da “incerteza”, tendo em conta a informação disponível. Sendo o nosso objectivo validar a precisão das inferências feitas e não classificá-las como inferências correctas/incorrectas. Neste seguimento, interessa sublinhar que no decorrer deste trabalho, não se procurará entrar na discussão ‘Inferência Clássica’ vs ‘Inferência Bayesiana’, mas sim aplicar as técnicas/procedimentos respectivos, num caso específico, confrontando as estimativas feitas.

Na opinião de vários autores, a forma mais fácil para se compreenderem os fundamentos da inferência bayesiana é confrontá-los com os da inferência clássica. Como tal, segue-se uma abordagem, ainda que generalista, aos dois paradigmas em estudo (clássico e bayesiano), iniciando-se com uma referência bastante sucinta à inferência clássica, por se suporem conhecidos os aspectos mais elementares e posteriormente, com maior enfoque, à inferência bayesiana.

1.2.1. INFERÊNCIA CLÁSSICA

De acordo com a bibliografia, autores como Karl Pearson, Ronald A. Fisher e Jerzy Neyman tiveram forte contributo para o predomínio da inferência clássica verificado há umas décadas atrás.

De acordo com Paulino *et al.* (2003, p. 2), a amostra, enquanto “conjunto de todas as observações possíveis ou concebíveis feitas em condições semelhantes”, assume um papel

determinante no quadro clássico, na medida em que o principal objectivo da inferência estatística é determinar, quando possível, quais as generalizações que podem ser feitas sobre a população, tendo por base os aspectos que a amostra, que desta foi recolhida, indicia.

Desta ideia, sobressai de imediato um aspecto fulcral: a variabilidade da amostra. Isto implica, na inferência clássica, que os dados em análise (a amostra) são apenas um dos muitos conjuntos passíveis de serem obtidos nas mesmas circunstâncias.

Assim, existindo essa incerteza quanto à amostra (possíveis conjuntos de dados), fará sentido que os dados estatísticos, $x = (x_1, \dots, x_n)^2$, onde n representa a dimensão da amostra e x é um elemento do espaço-amostra, \mathcal{X} , sejam aceites como variáveis aleatórias, $X = (X_1, \dots, X_n)$, com função de distribuição F_0 , a qual não é perfeitamente conhecida. Porém, dado algum conhecimento que se possa ter do fenómeno aleatório em estudo, podemos assumir uma família de distribuições, \mathcal{F} , onde cada elemento dessa família é um candidato a F_0 . A escolha desta família de distribuições pode resultar de uma complexa síntese de factores, os quais podem ser consultados em Paulino *et al.* (2003).

Neste seguimento, contendo cada modelo ‘quantidades’ observáveis e outras não directamente observáveis (aquelas que se procuram estimar), consideramos θ um vector de parâmetros a ser estimado, o qual está restrito a um conjunto Θ , designado como espaço-parâmetro.

Pelo que, sendo as distribuições, \mathcal{F} , representadas pelas respectivas densidades, o modelo estatístico pode ser representado, em função das densidades marginais das variáveis X_1, \dots, X_n , por

$$\mathcal{F} = \left\{ f(x|\theta) = \prod_{i=1}^n f_i(x_i|\theta) : \theta \in \Theta \right\}, x \in \mathcal{X}$$

para o caso das variáveis aleatórias $X = (X_1, \dots, X_n)$ serem independentes condicionalmente a θ .

Além da modelação e parametrização, existem outros procedimentos na inferência clássica que visam explorar quais as conclusões que do modelo, extraído da amostra, se podem inferir para a população. Tal como é referido em Paulino *et al.* (2003, p. 4), “os procedimentos clássicos, são julgados à luz da amostragem repetida”. Segundo este princípio, “os métodos estatísticos devem apreciar-se através do respectivo

² No caso da amostra ter dimensão 1, x denota um escalar (apenas uma observação). Nesta apresentação teórica, consideramos o caso geral, como tendo uma amostra (finita) de dimensão n , pelo que, quando nos referirmos a x ou outras componentes, estaremos sempre a pensar num vector e não apenas num escalar.

comportamento num indefinido de repetições – hipotéticas – efectuadas nas mesmas condições”. Deste modo, será, talvez, lícito afirmar que a interpretação frequencista de probabilidade assume-se como a ‘medida da incerteza’ que se pretende estudar.

Esta é uma das possíveis interpretações do conceito de probabilidade³ – interpretação frequencista – em que, considerando uma sucessão numerosa de provas/observações (idênticas e independentes), sustenta que o valor da probabilidade de um dado acontecimento pode ser traduzido pela frequência relativa do mesmo.

Ainda, numa fase cujo objectivo principal é tirar conclusões, tem particular interesse as que podem ser extraídas do verdadeiro valor dos parâmetros, pertencentes ao vector θ , que descrevem de forma adequada as condições em que se observam as variáveis aleatórias. Aqui surgem conceitos como ‘eficiência’ dos estimadores, intervalos de confiança, testes de hipóteses, entre outros ligados à estimação e ensaios de hipóteses, que apesar importantes, não serão aqui analisados, por serem passíveis de fácil consulta (com rigor e detalhe) numa vasta bibliografia referente à área da estatística clássica, como é exemplo Murteira (1990).

Finalmente, sendo esta breve alusão à inferência clássica um ponto de partida para a apresentação de aspectos relacionados com a inferência bayesiana, não obstante da não entrada em confronto directo dos dois paradigmas (clássico e bayesiano), será interessante referir duas opiniões contrárias sobre o “pensamento frequencista”.

Oliveira (*apud* Paulino *et al.*, 2003, p. 8) afirma que os bayesianos, em particular,

“... introduzem novos princípios (filosóficos) de inferência... A inferência clássica, agarrada à interpretação empirista das probabilidades, apenas tem de procurar critérios mais ajustados pois não perde o sentido do real. Nas outras inferências a escolha é a gosto num largo leque de princípios mas o real escapa-se”.

Já Jaynes (*apud* Paulino *et al.*, 2003, p. 8) apresenta uma opinião contraditória ao afirmar que

“... the traditional ‘frequentist’ methods which use only sampling distributions are usable and useful in many particularly simple, idealized problems; but they represent the most proscribed special cases of probability theory, because they presuppose conditions (independent repetitions of a ‘random’ experiment but no relevant priori information) that are hardly ever

³ Outras interpretações do conceito de probabilidade podem ser consultadas em Paulino *et al.* (2003).

met in real problems. This approach is quite inadequate for the current needs of science”.

1.2.2. INFERÊNCIA BAYESIANA

Apesar do impacto ter sido maior nas últimas décadas, as referências às ideias bayesianas e à sua aplicabilidade em termos estatísticos começaram a ganhar dimensão, na perspectiva de vários autores, com Harold Jeffreys, como é exemplo o trabalho publicado em 1939 “*Theory of Probability*”, em que, ao refutar a posição clássica, consegue dar um impacto ao paradigma bayesiano, avançando com soluções para problemas estatísticos que, até ao momento, persistiam sem solução. Tal facto, parece ter contribuído para que todo um grupo de autores fosse surgindo na defesa dos procedimentos bayesianos: Good (1950), Savage (1962), Lindley (1965), entre outros citados em vários trabalhos relacionados com a inferência bayesiana.

Sendo a inferência bayesiana um pilar para o(s) modelo(s) que nos propomos desenvolver nesta dissertação, reconhecemos a multiplicidade de factos que seria desejável referir. Porém, procurando evitar uma citação de conteúdos, que tornaria bastante extensa e complexa esta apresentação, optámos por fazer uma síntese dos principais enfoques do paradigma bayesiano, procurando adaptá-lo ao caso econométrico.

Sendo a ‘medida da incerteza’ feita por meio de probabilidades, na teoria bayesiana a interpretação deste conceito, associado à ocorrência de um determinado evento incerto, não é a interpretação frequentista, contrapondo com a inferência clássica. Os procedimentos da inferência bayesiana, ao descreverem probabilisticamente as ‘quantidades’ não observáveis, têm por base as regras básicas da teoria de probabilidades. Pelo que, será lícito afirmar que os métodos bayesianos são uma extensão dos métodos clássicos, divergindo de uma forma bastante sucinta no seguinte facto: sendo a incerteza expressa por probabilidades a essência dos métodos bayesianos, aplicando a essas probabilidades *a priori* o Teorema de Bayes, estas mudam, resultando em probabilidades *a posteriori*.

Considerando θ um vector de parâmetros a ser estimado, com $\theta \in \Theta$, se na inferência clássica este vector, embora desconhecido, é fixo, segundo os procedimentos bayesianos θ é um vector aleatório (não observável), pelo que se assume que o vector θ , sendo desconhecido, é incerto e toda essa incerteza é quantificada em termos de probabilidade (Koop, 2004).

Assim, poder-se-á afirmar que o conhecimento de θ , é dado por meio de uma função de probabilidade/função densidade de probabilidade (fdp), denominada por distribuição *a priori*, $h(\theta)$. Esta designação defendida pelo bayesianos, que acaba por distinguir a abordagem bayesiana das restantes, deriva do facto de se ter uma informação inicial⁴, anterior à experiência, traduzida formalmente por uma distribuição de probabilidade a qual, baseando-se na informação disponível, procura exprimir o grau de credibilidade atribuída pelo indivíduo, dependendo esta do próprio indivíduo, isto é, do que este julga ser mais ou menos provável, num contexto específico. Segundo a bibliografia respeitante ao tema, a determinação e consequente interpretação da distribuição *a priori*, é dos pontos mais questionáveis da teoria bayesiana e constitui um dos obstáculos à sua implementação. Esta interpretação subjectiva (específica de um dado problema e de um determinado investigador) da fdp., está igualmente patente na família de distribuições \mathcal{F} , pelo que os elementos $f(x|\theta) \in \mathcal{F}$ são também eles subjectivos.

“A discussão das distribuições a priori e da sua natureza subjectiva ilustra muito claramente alguns aspectos do confronto entre bayesianos e clássicos. Para os primeiros [Berger (1984)] a escolha subjectiva do modelo – família \mathcal{F} – traduz muitas vezes um uso mais drástico da informação a priori do que o emprego de uma distribuição a priori para o parâmetro θ , do modelo. Para os segundos, [Lehman (1983)], há uma importante diferença entre a modelação de \mathcal{F} e a modelação de $h(\theta)$ pois enquanto se dispõe de um conjunto de observações – (x_1, \dots, x_n) – geradas por um membro de \mathcal{F} que pode empregar-se para testar a forma da distribuição, o valor de θ é apenas uma observação da distribuição $h(\theta)$ ”. (Paulino et al., 2003, p.12)

Em suma, embora outros argumentos possam ser expostos, os autores apologistas da inferência bayesiana sublinham que um factor limitador da inferência clássica deve-se ao facto de não se atender à informação *a priori*, ou atenderem de forma informal. Pois, para estes, a informação *a priori* detida deve traduzir-se, formalmente, por uma distribuição de probabilidade para o vector aleatório, θ , não observável.

⁴ Alguns autores, como Jaynes (*apud* Paulino et al., 2003), denotam a designada distribuição por $h(\theta|I)$, para dar ênfase à ideia de que esta depende da informação inicial do indivíduo, I , o que fará sentido pois podem surgir alterações a informação inicial, I_0 vs I_1 , ou quando se compara soluções propostas por diferentes indivíduos, I_A vs I_B .

Na tentativa de esquematizar as principais ideias acima apresentadas, vejam-se as figuras seguintes, onde, confrontando os dois procedimentos em análise, clássico e bayesiano, ficam claras quais as divergências entre os mesmos.⁵

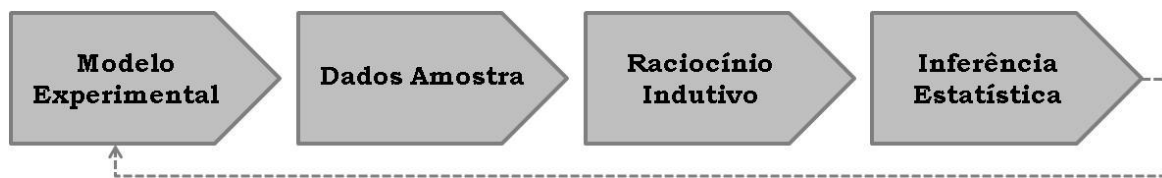


FIGURA 1 – Esquema ilustrativo dos procedimentos na Inferência Clássica

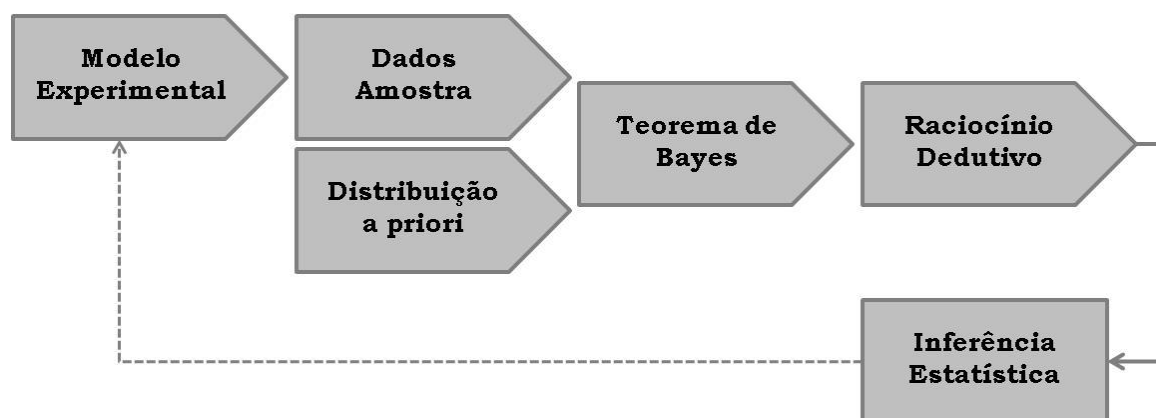


FIGURA 2 – Esquema ilustrativo dos procedimentos na Inferência Bayesiana

Depois de exposto, em linhas gerais, o paradigma bayesiano, será então necessário formalizar algumas das considerações acima feita.

Por ocupar um papel determinante na inferência bayesiana, é necessário fazer uma referência inicial ao Teorema de Bayes.

Será ainda de referir que toda a abordagem feita nas secções subsequentes terá por base a abordagem discreta. Isto é, a abordagem feita, na sequência do Teorema de Bayes, assume que o vector aleatório, θ , é constituído por variáveis aleatórias discretas com espaço-parâmetro, $\Theta = \{\theta_k : k = 1, \dots, n\}$, finito. Toda a exposição poderia ser feita para o caso contínuo, ou com um espaço de parâmetro, Θ , infinito, seguindo um raciocínio análogo.

⁵ Esquemas adaptados de Paulino *et al.*(2003).

1.2.3. TEOREMA DE BAYES

O Teorema de Bayes, atribuído a Thomas de Bayes⁶, é uma proposição sobre probabilidades condicionais, sendo que a sua interpretação e consequente utilização na inferência bayesiana têm gerado alguma controvérsia. Antes de enunciarmos o teorema em si, tenham-se em conta alguns requisitos, inicialmente definidos para dois acontecimentos, mas passíveis de ser generalizados, como veremos.

Observe-se que, em qualquer um das definições subsequentes partir-se-á de um espaço de probabilidade (Ω, \mathcal{A}, P) , onde:

- Ω é um espaço fundamental não vazio (que nas aplicações coincide, em regra denota o espaço amostral) cujos elementos ω , $\omega \in \Omega$, designamos de acontecimentos elementares e os subconjuntos A , $A \subseteq \Omega$, denotam os acontecimentos;
- \mathcal{A} é a família (álgebra ou σ -álgebra) dos acontecimentos dotados de probabilidade;
- P é a medida de probabilidade definida para os acontecimentos $A \subseteq \Omega$, $A \in \mathcal{A}$, em que $P(A)$ é a probabilidade do acontecimento A .

DEFINIÇÃO 1.1

Seja (Ω, \mathcal{A}, P) um espaço de probabilidades. Dados dois acontecimentos A e B , $(A, B \subseteq \Omega)$, a probabilidade de A se realizar, sabendo que se realizou B – a **probabilidade de A condicional a B** – denotada por $P(A|B)$, é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A, B)}{P(B)}, \quad P(B) > 0 \quad (1.1)$$

De forma análoga, definimos a **probabilidade de B condicional a A** , como:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A, B)}{P(A)}, \quad P(A) > 0 \quad (1.2)$$

Das igualdades (1.1) e (1.2), resulta que:

$$P(A, B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B) \quad (1.3)$$

⁶ Thomas Bayes nasceu em Londres por volta de 1702 e morreu a 17 de Abril de 1761 em Tunbridge Wells. Em 1719 ingressou, para estudar teologia e lógica, na Universidade de Edimburgo. No entanto os seus estudos continuariam numa universidade escocesa. Utilizando as probabilidades de forma intuitiva, estabeleceu as bases para a inferência estatística tornando-se conhecido por ter formulado o Teorema de Bayes. Apesar de não ter obras de matemática publicadas, Bayes foi eleito membro da Royal Society em 1742. Já após a sua morte, em 1763, Richard Price, amigo pessoal de Bayes publica a obra "An essay Towards Solving a problem in the Doctrine of Chances" a qual, segundo alguns autores, terá sido impulsionadora para o iniciar do estudo da abordagem bayesiana.

Manipulando algebricamente a equação (1.3), podemos deduzir o Teorema de Bayes, para dois acontecimentos.

TEOREMA 1.1 (Teorema de Bayes para dois acontecimentos)

Seja (Ω, \mathcal{A}, P) um espaço de probabilidades. Dados dois acontecimentos A e B , com $A, B \subseteq \Omega$ e $A, B \in \mathcal{A}$, tem-se:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \quad (1.4)$$

Se quisermos enunciar para o caso geral este resultado, tomemos para o efeito uma partição finita (ou infinita) de Ω dada por B_1, B_2, \dots, B_n , com $P(B_i) > 0$.

TEOREMA 1.2 (Teorema da Probabilidade Total)

Seja (Ω, \mathcal{A}, P) um espaço de probabilidades. Dado um acontecimento $A \subseteq \Omega$ e B_1, B_2, \dots, B_n , uma partição desse espaço, isto é:

i. $\bigcup_{i=1}^n B_i = \Omega$

ii. $B_i \cap B_j = \emptyset, \forall i \neq j$

Tem-se:

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i) \quad (1.5)$$

Consequentemente, tendo em conta as equações (1.4) e (1.5), podemos enunciar o Teorema de Bayes para o caso geral.

TEOREMA 1.3 (Teorema de Bayes – caso geral)

Seja (Ω, \mathcal{A}, P) um espaço de probabilidades. Dado um acontecimento $A \subseteq \Omega$ e B_1, B_2, \dots, B_n , uma partição desse espaço, temos:

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}, \quad k = 1, \dots, n \quad (1.6)$$

Esta abordagem é útil para os casos em que conhecemos as probabilidades dos acontecimentos B_1, B_2, \dots, B_n , e a probabilidade condicional de A dado B_i , com $i = 1, \dots, n$, mas não é conhecida, directamente, a probabilidade de A .

Estabelecendo uma junção de ideias entre a teoria de probabilidades e a econometria, neste caso tendo por base a inferência bayesiana (Econometria Bayesiana), pelo facto

desta trabalhar com modelos que dependem de parâmetros, os quais se pretendem estudar, o Teorema de Bayes pode ser apresentado numa nova perspectiva (Koop, 2004).

Assim, considerando y como sendo um vector (ou uma matriz) de dados e θ como sendo um vector (ou uma matriz) contendo os parâmetros do modelo que procuram explicar y , estamos interessados em conhecer o vector de parâmetros, θ , tomando como base o conhecimento de dados y . Pelo que, fazendo a junção de ideais acima referida, substituindo, na equação (1.4) (Teorema de Bayes), B por θ e A por y , obtemos:

$$p(\theta|y) = \frac{p(y|\theta).p(\theta)}{p(y)} \quad (1.7)$$

Sendo que, para o caso geral, se obtém:

$$p(\theta_k|y) = \frac{p(y|\theta_k).p(\theta_k)}{\sum_{i=1}^n p(y|\theta_k).p(\theta_k)}, \quad k = 1, \dots, n \quad (1.8)$$

A equação (1.7) evidencia a troca de uma “crença” a respeito do vector de parâmetros θ , expressa por $p(\theta)$, por um outra “crença” medida por $p(\theta|y)$. Com efeito, $p(\theta)$ pode ser interpretado como a distribuição de probabilidade de θ , antecedente à observação dos dados, e $p(\theta|y)$ a distribuição posterior à observação dos dados. Daí a designação de probabilidade *a priori* sobre θ para $p(\theta)$ e de probabilidade *a posteriori* para $p(\theta|y)$. Facto que vem clarificar a interpretação feita do Teorema de Bayes, conforme apresentada por vários autores, entre os quais Koop (2004).

Observe-se que todos os elementos da equação (1.7) dependem de θ à excepção do denominador da expressão, $p(y)$, referido por alguns autores como uma constante de normalização. Deste modo, sendo $p(\theta|y)$ precisamente o foco de estudo na econometria, é lícito afirmar que este depende na íntegra do produto de $p(y|\theta)$ por $p(\theta)$, isto é:

$$p(\theta|y) \propto p(y|\theta).p(\theta) \quad (1.9)$$

Ou seja, a densidade *a posteriori*, $p(\theta|y)$, é explicada por $p(y|\theta)$ denotando a função de verosimilhança – densidade dos dados, y , condicional aos parâmetros, θ , do modelo – e pela densidade *a priori*, $p(\theta)$. Ou seja, “a distribuição a posteriori incorpora, via Teorema de Bayes, toda a informação disponível sobre o parâmetro (informação inicial + informação da experiência ou da amostra)” (Paulino *et al.*, 2003, p.22), ou de uma forma mais simplista, diz-se que a distribuição *a posteriori* é proporcional ao produto da verosimilhança pela densidade *a priori* (Koop, 2004).

Observe-se ainda que, relativamente a cada um dos elementos de (1.9), temos:

- **a função de verossimilhança, $p(y|\theta)$** , é a fdp para os dados, y , condicional ao vector de parâmetros, θ . Teoricamente, uma função de verossimilhança vem no seguimento de uma probabilidade condicional considerada como uma função do seu argumento, mantendo o primeiro fixo. Sublinhe-se, com efeito, que na inferência bayesiana, contrariamente à inferência clássica, θ é tratado como uma variável aleatória, ao passo que os estimadores de θ , baseados na amostra, são tratados como fixos, condicionalmente aos dados observados. Esta expressão para a distribuição dos dados restrita aos parâmetros, pode ser descrita por distribuições simples sugeridas na teoria de probabilidades, ou por modelos de probabilidades para variáveis aleatórias mais complexos (Lancaster, 2004).
- **a densidade *a priori*, $p(\theta)$** , não depende dos dados observados, y , contendo apenas a informação disponível sobre θ . Esta expressão traduz uma probabilidade marginal reflectindo o conhecimento prévio sobre os parâmetros, pelo que expressa a incerteza sobre θ , anterior à observação dos dados. Daí a referência *a priori* como uma informação puramente subjectiva.

Alguns autores, como Paulino *et al.* (2003, p. 14 – 21) e Costa (2004, p. 3), apresentam analítica e graficamente alguns exemplos específicos do efeito desta distribuição de probabilidade sobre a distribuição *a posteriori*, sendo, em alguns casos, comparada a distribuição *a priori* com a verossimilhança e com a distribuição *a posteriori*, pondo em conflito várias situações. A título de exemplo, veja-se a figura seguinte, na qual a distribuição *a priori* contém informação suficiente para modificar substancialmente a distribuição *a posteriori*, gerando uma distribuição unimodal.

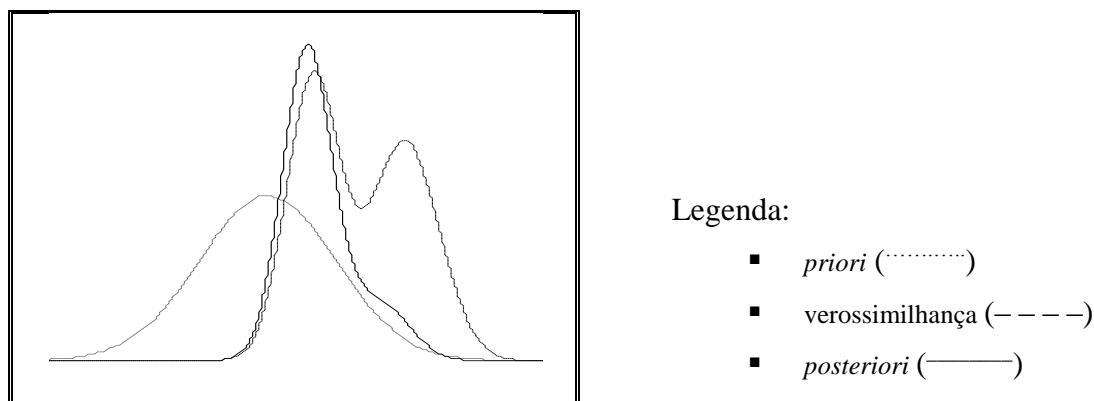


FIGURA 3 – Representação gráfica do efeito da distribuição *a priori*
(Adaptado de Costa, 2004, p. 3)

- a **densidade *a posteriori***, $p(\theta|y)$, expressa o conhecimento sobre os parâmetros posterior à observação dos dados, sendo traduzida por uma probabilidade condicional de uma variável aleatória, θ , dado um conjunto de observações, y .

Não sendo objectivo deste trabalho uma análise aprofundada e reconhecendo que as considerações feitas a respeito de cada uma das componentes foi bastante generalista e redutora, uma abordagem mais exaustiva pode ser consultada em Lancaster (2004), nomeadamente sobre alguns aspectos teóricos sobre verosimilhança (princípio da verosimilhança), tipos de *priori* e as formas mais comuns para a apresentação da distribuição *a posteriori*. No entanto, a nossa abordagem não cessa aqui, uma vez que na Secção 1.2.3. serão ainda tecidas algumas considerações sobre a distribuição *a priori* e é sobre a distribuição *a posteriori* que irá incidir parte da nossa abordagem teórica, nomeadamente na apresentação e análise de métodos de simulação de distribuições *a posteriori* (Secção 1.4).

Para finalizar esta análise, sem a querer tornar mais extensa, mas pela necessidade em relacionar todos os factos constantes nas secções anteriores, relativos à inferência bayesiana, procurando cimentar ideias pela generalização de raciocínio, observe-se que os parâmetros identificados podem referir-se a funções (ou até mesmo constantes). Para o efeito, supondo que se observa $X = x$, considere-se um qualquer elemento $f(x|\theta)$, pertencente à família de distribuições \mathcal{F} , e a distribuição *a priori* $h(\theta)$, conforme apresentado em Paulino *et al.* (2003), a equação (1.8) pode ser escrita na forma

$$h(\theta_k|x) = \frac{f(x|\theta_k) \cdot h(\theta_k)}{\sum_{i=1}^n f(x|\theta_k) \cdot h(\theta_k)}, \quad k = 1, \dots, n \quad (1.10)$$

onde, para cada $k = 1, \dots, n$, $h(\theta_k|x)$ denota a distribuição *a posteriori* de θ , depois de se observar x . Ou seja, tendo em conta a informação contida nos dados observados, x , a “crença” caracterizada por $h(\theta_k)$, passa a ser caracterizada por $h(\theta_k|x)$.

Ainda, se considerarmos a observação de uma amostra finita de dimensão m , (x_1, \dots, x_m) , a equação (1.10) vem:

$$h(\theta_k|x_1, \dots, x_m) = \frac{\prod_i f(x_i|\theta_k) \cdot h(\theta_k)}{\prod_i [\sum_{i=1}^n f(x_i|\theta_k) \cdot h(\theta_k)]}, \quad k = 1, \dots, n \quad (1.11)$$

Pelo que, simplificando a notação e sem perda de generalidade, pode finalmente escrever, pelas razões já expostas:

$$h(\theta|x) \propto f(x|\theta).h(\theta) \quad (1.12)$$

tal como feito em (1.9) para o caso particular.

1.2.4. REFERÊNCIA À DISTRIBUIÇÃO A PRIORI

Tal como já referido, na abordagem bayesiana, os parâmetros são variáveis aleatórias nos quais é passível de ser incorporada uma informação adicional, envolvendo elementos subjectivos. A quantificação desses elementos/crenças probabilísticas *a priori*, que se caracteriza pela sua subjectividade, é realizada por meio de distribuições de probabilidade, podendo estas ser informativas ou não informativas, as quais têm repercussões na distribuição *a posteriori*. Assim, quando se tem informação *a priori* sobre os parâmetros, esta pode ser expressa utilizando distribuições *a priori* informativas, cuja fonte é objectiva (dados históricos, por exemplo). Porém, caso não exista uma opinião consistente sobre os parâmetros em estudo, esse ‘desconhecimento’ *a priori* pode ser expresso através de distribuições *a priori* não-informativas.

É na quantificação probabilística da subjectividade acima referida que reside a principal dificuldade, a qual compreende conceitos ‘estranhos’ à inferência clássica.

No que respeita ao processo de especificação de probabilidades subjectivas⁷, esta é transformada numa distribuição *a priori* que será usada para prosseguir com os pressupostos da inferência bayesiana. Tal como descrito em Paulino *et al.* (2003, p. 72), “...o método mais usado consiste na adopção de uma forma distribucional conveniente e consentânea (...) destacando-se neste papel a denominada família das distribuições conjugadas⁸...”.

Com efeito, um procedimento seguido é a especificação através de formas funcionais de densidades paramétricas, cujos parâmetros são escolhidos de modo subjectivo, com base na informação disponível. Uma das metodologias possível consiste em escolher a forma funcional da distribuição *a priori* por forma a que ambas as distribuições, *a priori* e *a posteriori*, pertençam à mesma família de distribuições, designadas de famílias de distribuições conjugadas.

⁷ Uma probabilidade subjectiva, opondo-se ao conceito lógico de probabilidade, é uma medida de um grau de crença pessoal específico do indivíduo, logo é, regra geral, diferenciada para cada observador.

⁸ Dada uma família de distribuições amostrais (observáveis), $\mathcal{F} = \{f(y|\theta) : \theta \in \Theta\}$, uma classe, \mathcal{C} , é uma família de distribuições conjugadas, com respeito a \mathcal{F} , se $\forall f \in \mathcal{F}$, $p(\theta) \in \mathcal{C}$ se tem $p(\theta|y) \in \mathcal{C}$.

Não sendo objectivo a exploração deste tema no nosso trabalho, para uma análise detalhada sobre distribuições *a priori* subjectivas e distribuições *a priori* conjugadas, veja-se Paulino *et al.* (2003).

Fica, desta forma claro que, sendo propósito seguir a abordagem bayesiana, perante uma informação *a priori* (mais ou menos substancial) a tentativa será traduzi-la numa distribuição *a priori* subjectiva. Contudo, não existindo uma informação palpável (informação objectiva), ou sendo o conhecimento *a priori* relativamente à informação amostral pouco significativo, Paulino *et al.* (2003) fazem uma exploração detalhada sobre os principais métodos de geração de distribuições (conduzindo a distribuições as distribuições *a priori* minimamente informativas em algum sentido), na bibliografia designadas como distribuições não-informativas, a que vagamente já nos referimos. De entre os principais métodos, destacam o *Método de Bayes-Laplace*, o *Método de Jeffreys* e o *Método de Box-Tiao*. Não sendo, igualmente, uma finalidade de estudo no nosso trabalho, para uma leitura sobre o tema veja a síntese apresentada em Paulino *et al.* (2003), os quais remetem para várias outras referências bibliográficas, dada a imensurabilidade de aspectos em questão.

Finalmente, observe-se que, por não apresentar a subjectividade na escolha da distribuição, os parâmetros estimados usando *prioris* não-informativas, apresentarão valores mais próximos comparativamente aos estimados pelo processo clássico. Contudo, apesar de amplamente utilizadas, no caso do vector de parâmetros ser não limitado, as *prioris* não-informativas podem levantar problemas computacionais na obtenção das distribuições *a posteriori*. A solução proposta na literatura para este problema passa por encontrar *prioris* não-informativas que são invariantes a parametrizações.

1.3. CADEIAS DE MARKOV EM TEMPO DISCRETO

Para a introdução dos métodos de Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo* – MCMC) e respectivos algoritmos de simulação, com vista à sua melhor compreensão e ao modo como podem ser utilizados na inferência bayesiana, fará sentido uma apresentação prévia de alguns conceitos básicos, bem como a referência a alguns resultados teóricos alusivos a cadeias de Markov⁹. É isso que se pretende fazer nesta secção.

⁹ Andrei Markov foi um matemático de nacionalidade russa. Nascido em 1856, foi professor em São Petersburgo desde 1886, tendo desenvolvido trabalhos significativos nas áreas da análise numérica, teoria da aproximação e convergência de séries. Porém, os estudos desenvolvidos sobre as designadas *Cadeias de Markov* constituem a sua contribuição mais importante, estabelecendo a teoria dos processos estocásticos.

Porém, estamos cientes de que todo o encadeamento de ideias seguinte, poderia ser sintetizado num parágrafo antecedente à apresentação dos métodos MCMC. Todavia, a sua compreensão seria, certamente, redutora e incompleta. Pelo que tendo sido imprescindível um estudo dos resultados que sustentam estes métodos, pareceu-nos pertinente que toda essa informação constasse neste trabalho, contrapondo com a hipótese de apresentação única e exclusiva da síntese de ideias que é apresentada no final desta secção.

Acrescente-se que, não sendo o nosso objectivo citar resultados concernentes à teoria das cadeias de Markov, fez-se uma análise prévia das definições e resultados teóricos implícitos na aplicação dos métodos MCMC, sendo que apenas esses (e outros necessários à compreensão dos mesmos), serão os focados neste trabalho.

Note-se que esta exposição terá por base o caso discreto – *Cadeias de Markov em Tempo Discreto* – uma vez que nos propomos fazer, nas secções subsequentes, uma análise de séries temporais. Porém, sublinhe-se que a abordagem para o caso contínuo seria análoga, havendo o cuidado de adaptar definições e resultados.¹⁰

Precedente à formalização de algumas noções e resultados, apresente-se, em linhas gerais, a ideia do conceito fulcral desta secção – *Cadeia de Markov*.

Tal como descrito em Esquivel (2009), onde se pode consultar um exemplo concreto de aplicabilidade, considerando o estudo de um fenómeno em tempo discreto, considere-se uma população onde cada elemento pode, num dado instante, encontrar-se num determinado estado; estados esses que podem pertencer a um conjunto finito, ou infinito numerável. Uma das pretensões, poderá ser estipular a forma como a classificação dos elementos da população evolui, segundo o estado em que se encontra. Para o efeito, considere-se uma variável aleatória, X_n , definida num conjunto Ω (população), tomando valores no espaço de estados S , que, num dado momento, $n \in \mathbb{N}_0$, nós dá a situação de cada elemento da população. Entre outras possibilidades, a forma como se altera a situação da população ao longo do tempo pode ser especificada supondo conhecida, para cada $i, j \in S$ e para $n \in \mathbb{N}_0$, a probabilidade de que na data $n + 1$ se estará no estado j , sabendo que na data imediatamente anterior, n , se está no estado i , isto é:

$$p_{ij}(n) := P(X_{n+1} = j \mid X_n = i)$$

¹⁰ A análise teórica a cadeias de Markov, considerando um espaço de estados geral, pode ser consultada, entre outras referências, em Paulino *et al.*, (2003, cap.8).

Observe-se que, por vezes, a data em que nos encontramos, n , não é determinante na lei de evolução das probabilidades $p_{ij}(n)$, pelo que, simplificando a notação, podemos escrever $p_{ij}(n) \equiv p_{ij}$.¹¹

A ideia da forma como as transições entre estados se opera será algo determinante e de particular interesse. Mediante um número significativo de observações, é possível ter essa ideia a qual pode ser sistematizada numa matriz, \mathcal{P} , – matriz de transição entre estados (a um passo) da Cadeia de Markov – onde cada elemento, p_{ij} , representa a probabilidade de se estar, numa dada data, no estado j , sabendo que na data imediatamente anterior, se estava no estado i . Pelo que, para cada linha, i , da matriz \mathcal{P} , se tem $\sum_{j \in S} p_{ij} = 1$ e $p_{ij} \geq 0$.¹²

Finalmente, considerando que é observável a distribuição inicial da população pelos estados do conjunto S , sob certas condições, o comportamento a longo prazo da sucessão de variáveis aleatórias, $(X_n)_{n \in \mathbb{N}_0}$, é determinado pela matriz de transição, \mathcal{P} , e pela distribuição inicial. Com efeito, se o nosso objectivo é construir uma amostra de uma dada distribuição, a qual não é possível fazer directamente, admitimos ser possível construir uma cadeia, cuja distribuição de equilíbrio seja a da amostra pretendida, pelo que se se simularem estados por um longo período de tempo, a fim de garantir o equilíbrio pretendido, estes podem ser usados para o conhecimento das características da distribuição de interesse (Paulino *et al.*, 2003, p. 306). O desafio que se impõe, resume-se, então, ao modo como construir uma cadeia com uma distribuição de equilíbrio específica.

É precisamente a definição/descrição de propriedades dessa cadeia de variáveis aleatórias – Cadeia de Markov – que nos propomos analisar de seguida, sendo a mesma feita em três partes. Na primeira, constam as noções gerais relativas a cadeias de Markov, bem como as características de interesse para a nossa apresentação. Na segunda constam alguns requisitos (notações e definições) relativos a distribuições das cadeias de Markov e são apresentados os principais resultados de interesse (condições de existência e unicidade de distribuições estacionárias, bem como a respectiva convergência). Finalmente, são tecidas algumas conclusões estabelecendo uma ponte entre os factos referidos concernentes às cadeias de Markov e a secção subsequente – MCMC.

Convém ainda referir que, não se tendo seguido a nomenclatura, nem a estrutura de nenhum autor específico, dado que, como já foi referido, o interesse na teoria das cadeias

¹¹ Como veremos mais à frente, estes são os casos de particular interesse no nosso estudo, o que se irá definir como uma Cadeia de Markov Homogénea, daí o uso, daqui em diante, da notação simplificada.

¹² Uma matriz nestas condições, diz-se uma *matriz estocástica*.

de Markov é pontual (apenas serão focados aspectos considerados relevantes), procurou-se construir uma estrutura própria e usar uma notação resultante do contributo de vários autores consultados para o efeito, como: Paulino *et al.* (2003), Esquível (2009) e, em particular, Hoel *et al.* (1987) e Grimmett e Stirzaker (1992). Sugerem-se estes últimos para uma leitura mais aprofundada e consulta de demonstrações de resultados constantes nesta secção.

1.3.1. NOÇÕES GERAIS EM CADEIAS DE MARKOV

Um requisito prévio é a noção de Processo Estocástico num espaço de estados, S . Sem perda de generalidade, considerar-se-á que o conjunto S é numerável e finito, com dimensão N , $S = \{1, 2, \dots, N\}$.

DEFINIÇÃO 1.2

Seja S um conjunto de estados, onde cada elemento é chamado de estado. Um **processo estocástico** em tempo discreto em S , é uma sucessão de variáveis aleatórias, $(X_n)_{n \in \mathbb{N}_0}$, tal que X_n assume valores em S .

A transição entre estados (definida pela *função de transição*) e a *distribuição inicial*, são os outros conceitos implícitos na definição de *Cadeia de Markov*.

DEFINIÇÃO 1.3

Dado o processo $(X_n)_{n \in \mathbb{N}_0}$, define-se a **função de transição** entre estados (a um passo), do estado i para o estado j , como a aplicação $p: S \times S \rightarrow [0,1]$ definida por:

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) \quad (1.13)$$

onde p satisfaz as seguintes condições:

- (i) $\forall i, j \in S, p_{ij} \geq 0$.
- (ii) $\forall i \in S, \sum_{j \in S} p_{ij} = 1$.

O conjunto de elementos p_{ij} pode ser organizado numa matriz, \mathcal{P} , denominada *matriz de transição*, sendo que, dados os estados $i, j \in S$, se tem:

$$\mathcal{P} = [p_{ij}]_{i,j \in S}$$

DEFINIÇÃO 1.4

Dado o processo $(X_n)_{n \in \mathbb{N}_0}$, define-se a **distribuição inicial**, como a aplicação $\pi^{(0)}: S \rightarrow [0,1]$ definida por:

$$\pi_i^{(0)} = p_i = P(X_n = i)$$

onde $\pi^{(0)}$ satisfaz as seguintes condições:

- (i) $\forall i \in S, p_i \geq 0.$
- (ii) $\sum_{i \in S} p_i = 1.$

Sintetizando a informação das definições precedentes, podemos definir os processos com particular interesse no nosso estudo: processos que, dado o estado presente, os estados passados não têm influência no estado futuro, ou seja, dado o presente, (X_n) , o futuro, (X_{n+1}) , e o passado, $(X_0, X_1, \dots, X_{n-1})$ são independentes – *Propriedade de Markov* – equação (1.13).

DEFINIÇÃO 1.5

Um processo estocástico em tempo discreto em S , $(X_n)_{n \in \mathbb{N}_0}$, com distribuição inicial $\pi^{(0)}$ e matriz de transição $\mathcal{P} = [p_{ij}]_{i,j \in S}$, diz-se uma **Cadeia de Markov** (finita e em tempo discreto) com espaço de estados S se:

- (i) $\forall i \in S, P(X_0 = i) = p_i = \pi_i^{(0)}.$
- (ii) Dado $n \in \mathbb{N}_0, \forall i, j \in S, P(X_{n+1} = j | X_n = i) = p_{ij}.$
- (iii) Dado $n \in \mathbb{N}_0, \forall i_0, i_1, \dots, i_{n-1}, i, j \in S,$
 $P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i) = p_{ij}$

A título ilustrativo, apresenta-se em anexo um exemplo (Exemplo 1.1. – Anexo A), em contexto real, do conceito de Cadeia de Markov.

Uma característica de interesse para o nosso estudo, tal como já foi referido, é quando a probabilidade indicada na equação (1.13) é independente de n , neste case diz-se que a cadeia é *homogénea*.

DEFINIÇÃO 1.6

Uma cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, em S (com espaço de estados S), diz-se **homogénea** se, dado $n \in \mathbb{N}_0$,

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = p_{ij}, \quad \forall i, j \in S$$

Daqui em diante, salvo observação em contrário, quando consideramos uma *Cadeia de Markov*, $(X_n)_{n \in \mathbb{N}_0}$, consideramos sempre que estamos a trabalhar com cadeias homogéneas, isto é, as probabilidades de transição de um estado para outro são invariantes.

Semelhante à forma como se definiu a função de transição para um passo, podemos definir a *função de transição a m-passos*.

DEFINIÇÃO 1.7

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S . Dado $m \in \mathbb{N}_0$ e $i, j \in S$, define-se a **função de transição entre estados em m-passos**, como a aplicação $p^{(m)}: S \times S \rightarrow [0,1]$ definida por:

$$p_{ij}^{(m)} = P(X_{n+m} = j \mid X_n = i) = P(X_m = j \mid X_0 = i)$$

Em particular:

- (i) $p_{ij}^{(0)} = P(X_0 = j \mid X_0 = i) = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases}$
- (ii) $p_{ij}^{(1)} = p_{ij}$

O conjunto de elementos $p_{ij}^{(m)}$, à semelhança para a transição a um passo, pode ser organizado numa matriz, $\mathcal{P}^{(m)}$, denominada *matriz de transição de ordem m*, sendo que, dados os estados $i, j \in S$, se tem:

$$\mathcal{P}^{(m)} = \left[p_{ij}^{(m)} \right]_{i,j \in S}$$

Tendo em conta a definição, observe-se que:

- para o caso $m = 0$, a matriz

$$\mathcal{P}^{(0)} = \left[p_{ij}^{(0)} \right]_{i,j \in S} = Id_S,$$

onde Id_S denota a matriz identidade para quaisquer estados $i, j \in S$;

- o caso $m = 1$, corresponde à Definição 1.3;
- para o caso $m = 2$, podemos definir a matriz

$$\mathcal{P}^{(2)} = \mathcal{P} \cdot \mathcal{P} = \left[p_{ij}^{(2)} \right]_{i,j \in S},$$

onde, tendo em conta o produto de matrizes usual, podemos escrever:

$$p_{ij}^{(2)} = \sum_{k \in S} p_{ik} \cdot p_{kj}$$

- generalizando o raciocínio, podemos definir a matriz

$$\mathcal{P}^{(m)} = \mathcal{P} \cdot \mathcal{P}^{(m-1)},$$

onde, recorrendo ao produto usual de matrizes, temos:

$$p_{ij}^{(m)} = \sum_{k \in S} p_{ik} \cdot p_{kj}^{m-1}$$

Além da homogeneidade, os métodos MCMC requerem, ainda que uma cadeia seja *irredutível*, isto é, cada estado poderá se atingido, a partir de qualquer outro, num número finito de iterações. A apresentação formal deste conceito requer a clarificação da relação “*i comunica com j*”, tendo em conta a noção de *acessibilidade* entre estados.

DEFINIÇÃO 1.8

Dado o espaço de estados S , um estado $j \in S$ diz-se **acessível** a partir do estado $i \in S$ (denotamos $i \rightarrow j$) se para algum $m \in \mathbb{N}_0$ se tem $p_{ij}^{(m)} > 0$.

DEFINIÇÃO 1.9

Dado o espaço de estados S , diz-se que um estado $i \in S$ **comunica** com o estado $j \in S$ (denotamos $i \leftrightarrow j$) se $i \rightarrow j$ e $j \rightarrow i$.

Observe-se que a relação “ \leftrightarrow ” é uma relação de equivalência (reflexiva, simétrica e transitiva). A demonstração deste resultado pode ser consultada em Esquível (2009).

Deste modo, podemos então definir *cadeia irredutível*.

DEFINIÇÃO 1.10

Uma Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, em S , diz-se **irredutível** se todos os estados, em S , comunicam entre si.

Procurando clarificar o conceito de irredutibilidade de uma cadeia, veja-se o Exemplo 1.2. apresentado em Anexo (Anexo A).

Se denotarmos por C_i a classe de todos os estado que comunicam com o estado $i \in S$, isto é, $C_i = \{k \in S : i \leftrightarrow k\}$, observe-se que, sendo a relação “ \leftrightarrow ” uma relação de equivalência, então numa cadeia irredutível todos os estados se encontram na mesma classe.

Outra característica que os métodos MCMC requerem está relacionada com a periodicidade da cadeia. Assim, vejamos previamente o que se entende por *período do estado*.

DEFINIÇÃO 1.11

Dado o espaço de estados S , um estado $i \in S$ tem **período** d se, dado $m \in \mathbb{N}_0$, se tem $p_{ii}^{(m)} = 0$, a menos que m seja divisível por d e d seja o maior inteiro nas condições enunciadas.

Ainda, $i \in S$ diz-se um estado **periódico** se $d > 1$. Caso $d = 1$, então dizemos que o estado i é **aperiódico**.

Neste seguimento, podemos definir o *período de uma cadeia*, bem como *cadeia aperiódica*.

DEFINIÇÃO 1.12

Dada uma Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, irredutível, com espaço de estados S , diz-se que $(X_n)_{n \in \mathbb{N}_0}$ tem período d se todos os estados têm período d .

Com efeito, se a cadeia tem período $d=1$, isto é, se todos os estados são aperiódicos, então dizemos que a cadeia é aperiódica.

Segundo Ross (*apud* Paulino *et al.*, 2003, p. 307), dado que a periodicidade é uma propriedade da classe, é condição suficiente, para que uma Cadeia de Markov irredutível seja aperiódica, que um dos seus estados seja aperiódico.

O valor da probabilidade de uma cadeia, iniciada em i , passar pelo estado j em algum momento finito, permite-nos distinguir *estados transientes* de *estados recorrentes*. Para o efeito, defina-se previamente a variável aleatória, τ_i , que representa o tempo da primeira passagem pelo estado $i \in S$.

DEFINIÇÃO 1.13

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S com função de transição p . Dados um estado arbitrário $i \in S$, define-se, caso exista, o **instante da primeira passagem** pelo estado i como a variável aleatória τ_i definida por:

$$\tau_i = \min \{n \in \mathbb{N}_0 : X_n = i\}$$

No caso de se ter $X_n \neq i$, para qualquer $n \in \mathbb{N}_0$, então dizemos que $\tau_i = \infty$.

Observe-se que, em particular, se o estado inicial da cadeia é i , então τ_i representa o tempo do primeiro regresso a i .

Com efeito, denotando por $\alpha_{ij} = P(\tau_j < \infty \mid X_0 = i)$, defina-se *estado transiente* e *estado recorrente*.

DEFINIÇÃO 1.14

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S . Um estado $i \in S$ diz-se:

- (i.) **transiente** se $\alpha_{ii} = P(\tau_i < \infty \mid X_0 = i) < 1$, isto é, condicionalmente a ter iniciado em i , há uma probabilidade positiva da cadeia nunca mais regressar a i .
- (ii.) **recorrente** se $\alpha_{ii} = P(\tau_i < \infty \mid X_0 = i) = 1$, isto é, condicionalmente a ter iniciado em i , a cadeia regressa garantidamente a esse estado.
- (iii.) **recorrente positivo** se o tempo médio da cadeia regressar ao estado i , condicionalmente a ter saído de i , é finito.

Observe-se que:

(i) se um estado, i , for transiente, então a cadeia iniciada em i , tem uma probabilidade positiva de nunca mais regressar a esse estado, pois:

$$P(\tau_i = \infty \mid X_0 = i) = 1 - P(\tau_i < \infty \mid X_0 = i) = 1 - \alpha_{ii} > 0$$

(ii) todo o estado recorrente verifica $p_{ii} = 1$, sendo que um estado que verifica esta condição, ($p_{ii} = 1$), diz-se *absorvente*.

Neste seguimento, a última característica requerida pelos métodos MCMC, relativamente às cadeias de Markov, resulta destas duas noções – *Cadeia Ergódica*.

DEFINIÇÃO 1.15

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S . Um estado $i \in S$ diz-se *ergódico* se é aperiódico e recorrente positivo.

Uma Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, em S , diz-se **ergódica** se é irredutível e aperiódica com todos os seus estados recorrentes positivos.

1.3.2. DISTRIBUIÇÃO ESTACIONÁRIA: CONDIÇÃO DE EXISTÊNCIA, UNICIDADE E CONVERGÊNCIA

As definições anteriores permitem estabelecer resultados, os quais são a base dos métodos MCMC. Apresentam-se de seguida as definições relativas às *distribuições de uma Cadeia de Markov e distribuição estacionária*.

TEOREMA 1.4 (Distribuição conjunta de $(X_n)_{n \in \mathbb{N}_0}$)

Sejam $n \in \mathbb{N}_0$ e $i_0, \dots, i_n \in S$. Dada uma Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, em S com função de transição p e distribuição inicial $\pi^{(0)}$, tem-se que:

$$P(X_0 = i_0, \dots, X_n = i_n) = \pi_{i_0}^{(0)} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}$$

Tendo em conta que:

$$\begin{aligned} p_j^{(n)} = P(X_n = j) &= \sum_{i \in S} P(X_0 = i, X_n = j) \\ &= \sum_{i \in S} P(X_0 = i) \cdot P(X_n = j \mid X_0 = i) \\ &= \left(\sum_{i \in S} P(X_{n-1} = i) \cdot P(X_n = j \mid X_{n-1} = i) \right) \end{aligned}$$

É possível enunciar ao seguinte resultado:

TEOREMA 1.5 (Distribuição marginal de $(X_n)_{n \in \mathbb{N}_0}$)

Sejam $n \in \mathbb{N}_0$ e $i, j \in S$. Dada $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S , com função de transição p e distribuição inicial $\pi^{(0)}$, tem-se que:

$$\pi_j^{(n)} = P(X_n = j) = \sum_{i \in S} \pi_i^{(0)} p_{ij}^{(n)}$$

Ou, alternativamente, podemos escrever:

$$\pi_j^{(n)} = P(X_n = j) = \sum_{i \in S} \pi_i^{(n-1)} p_{ij}$$

Observe-se que, para cada $n \in \mathbb{N}_0$, é possível pensar em $\pi^{(n)}$ como um vector em S . Pelo que, dada uma Cadeia de Markov com matriz de transição a (\cdot) -passos, $\mathcal{P}^{(\cdot)}$, os

resultados supramencionados podem ser reescritos, em notação matricial, da seguinte modo:

$$(i) \quad \mathcal{P}^{(n+m)} = \mathcal{P}^{(n)}\mathcal{P}^{(m)}, \quad \forall n, m \in \mathbb{N}_0$$

$$(ii) \quad \pi^{(n)} = \pi^{(0)}\mathcal{P}^{(n)}, \quad \forall n \in \mathbb{N}_0$$

$$(iii) \quad \pi^{(n)} = \pi^{(n-1)}\mathcal{P}, \quad \forall n \in \mathbb{N}_0$$

DEFINIÇÃO 1.16

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S com função de transição p . Diz-se que a função $\pi: S \rightarrow [0,1]$ é uma **distribuição estacionária** para $(X_n)_{n \in \mathbb{N}_0}$ se:

$$(i) \quad \forall i \in S, \quad \pi_i \geq 0.$$

$$(ii) \quad \sum_{i \in S} \pi_i = 1.$$

$$(iii) \quad \forall j \in S, \quad \sum_{i \in S} \pi_i p_{ij} = \pi_j. \text{ (Condição de equilíbrio)}$$

À semelhança do feito anteriormente, dada uma Cadeia de Markov ergódica em S , com matriz de transição \mathcal{P} , sendo S finito, é possível pensar em π como um vector em S . Nesse caso, tem-se, em notação matricial, que:

- $\pi = \pi\mathcal{P}$
- $\pi = \pi\mathcal{P}^{(n)}, \quad \forall n \in \mathbb{N}_0$

Tal como referido em Paulino *et al.* (2003, p. 308), no que se segue interessa-nos considerar sempre cadeias de Markov ergódicas. Neste caso pode mostra-se que, supondo a existência de uma distribuição estacionária, π , tal que $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$, para qualquer $j \in S$, temos:

- $\lim_{n \rightarrow \infty} \pi_j^{(n)} = \lim_{n \rightarrow \infty} P(X_n = j) = \pi_j, \quad \forall j \in S.$
- π é a única distribuição estacionária da cadeia $(X_n)_{n \in \mathbb{N}_0}$.

Ou seja, se se quiser “amostrar” de uma distribuição que atribui ao estado j probabilidade $\pi_j > 0$, tem de se construir uma Cadeia de Markov ergódica com matriz de transição \mathcal{P} , tal que a *condição de equilíbrio* é satisfeita Paulino *et al.* (2003, p. 308).

Seguem-se os resultados de interesse que garantem a existência (e unicidade) de distribuições estacionárias e respectiva convergência.

TEOREMA 1.6 (Adaptado de Grimmett e Stirzaker, 1992)

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov ergódica em S . Então, existe uma única distribuição estacionária, π , dada por:

$$\pi_i = \frac{1}{\mathbb{E}(\tau_i | X_0 = i)}, \forall i \in S$$

A expressão $\mathbb{E}(\tau_i | X_0 = i)$ denota o valor esperado do tempo da cadeia regressar ao estado inicial, isto é, do primeiro regresso ao estado i , sabendo que a cadeia se inicia em i .

TEOREMA 1.7 (Adaptado de Hoel, *et al*, 1987 e Tierney, 1994)

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov ergódica em S , com distribuição estacionária π e função de transição p .

Então:

- (i) $\sum_{i \in S} \pi_i p_{ij}^{(n)} = \pi_j, \forall j \in S$.
- (ii) Se $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j, j \in S$, independentemente da distribuição X_0 , temos:

$$\lim_{n \rightarrow \infty} \pi_j^{(n)} = \lim_{n \rightarrow \infty} P(X_n = j) = \pi_j$$

O que equivale a escrever:

$$\|\pi^{(n)} - \pi\| \xrightarrow{n} 0. \quad ^{13}$$

A demonstração de cada um destes resultados, por poder ser consultada na referência bibliográfica indicada, não será objecto de análise neste trabalho.

Para uma melhor compreensão dos aspectos em causa, veja-se o exemplo numérico (Exemplo 1.3.) apresentado no Anexo A.

Paulino *et al.* (2003, p. 308), acrescentam ainda uma outra propriedade com interesse – *reversibilidade*. Com efeito, para terminar, apresenta-se a definição de *Cadeia Reversível*.

¹³ Dada uma sucessão, $\pi, \pi^{(1)}, \dots, \pi^{(n)}$ de distribuições de probabilidade em S , dizemos que $\pi^{(n)}$ converge em variação total para π , quando $n \rightarrow \infty$, se a distância da variação total entre $\pi^{(n)}$ e π , $d(\pi^{(n)}, \pi)$, tende para zero, com $d(\pi^{(n)}, \pi) := \frac{1}{2} \sum_{i \in S} |\pi_i^{(n)} - \pi_i|$.

DEFINIÇÃO 1.17

Seja $(X_n)_{n \in \mathbb{N}_0}$ uma Cadeia de Markov em S . Diz-se que uma distribuição de probabilidade $\pi: S \rightarrow [0,1]$ é uma **distribuição reversível** para $(X_n)_{n \in \mathbb{N}_0}$ se:

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i, j \in S \ (i \neq j)$$

A cadeia $(X_n)_{n \in \mathbb{N}_0}$ diz-se **reversível** se existe uma distribuição reversível para a mesma.

Com efeito, tendo em conta a definição, temos:

$$\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \sum_{i \in S} p_{ji} = \pi_j$$

Então, nas condições enunciadas, podemos afirmar que a distribuição reversível, π , é uma distribuição estacionária para $(X_n)_{n \in \mathbb{N}_0}$. Tal facto vem directamente da definição. A este propósito, veja-se o Exemplo 1.4. apresentado em anexo (Anexo A).

Um dos métodos MCMC que se irá analisar na secção seguinte – algoritmo de *Metropolis-Hasting* – é um dos algoritmos que permitem construir matrizes de transição com esta propriedade.

1.3.3. DAS CADEIAS DE MARKOV AOS MÉTODOS MCMC

Com esta secção pretendeu-se apresentar as características da Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, requeridas pelos métodos MCMC:

- homogénea, isto é, as probabilidades de transição de um estado para outro são invariantes (independente de n);
- ergódica, isto é, a cadeia é irredutível (todos os estados comunicam entre si) e aperiódica com todos os seus estados recorrentes positivo (o tempo médio da cadeia regressar ao um determinada estado i , condicionalmente a ter saído de i , é finito)

Verificadas estas características, a questão é como os valores iniciais influenciam o comportamento da cadeia e para que distribuição esta tende (caso exista tendência). A ideia fulcral dos resultados apresentados, é que conforme o número de iterações aumenta (estatisticamente, está implícita a Lei dos Grandes Números), a cadeia gradualmente converge para uma distribuição de equilíbrio.

Assim, os métodos MCMC, apresentados na secção seguinte, usam processos estocásticos Markovianos com matriz de transição ergódica (no caso da cadeia ser

ergódica, a matriz de transição entre estados diz-se ergódica) para gerar, via simulação de Monte Carlo, amostras da distribuição de interesse, uma vez que a distribuição estacionária da matriz de transição converge para a distribuição dos parâmetros de interesse.

Os algoritmos de *Gibbs* e de *Metropolis-Hasting* são exemplos de processos que podem ser usados para encontrar a referida matriz de transição. Como resultado deste processo, obtém-se, assim, uma amostra da distribuição de interesse, a qual é utilizada para obtermos estimativas dos parâmetros da distribuição.

1.4. MÉTODOS DE SIMULAÇÃO ESTOCÁSTICA

A eficácia da análise econométrica depende, em parte, do grau de precisão em que os modelos estimados explicam/prevêem o comportamento das variáveis em estudo. Nem sempre sendo possível a obtenção de soluções via analítica, recorrendo a ferramentas matemáticas, impõe-se, em situações mais complexas, a utilização dos chamados métodos de simulação. De um modo sucinto, estes métodos consistem na implementação (a nível computacional), de algoritmos correspondentes à situação “real” que se deseja simular.

1.4.1. INTRODUÇÃO AOS MÉTODOS DE SIMULAÇÃO

A simulação pode revelar-se uma ferramenta bastante positiva, na medida em que permite fazer inferências, por meio de “experiências” que muitas vezes são passíveis de serem concretizadas em cenários reais, sobre aspectos de interesse. Com efeito, uma simulação desajustada pode comprometer todo o processo de modelação, o qual deve seguir uma sequência lógica, desde a identificação do problema em estudo, às conclusões do mesmo, passando, entre outras etapas cruciais, pela identificação e introdução de variáveis associadas ao problema, construção e consequente avaliação do modelo, mediante a realização de várias experiências as quais podem conduzir à necessidade de reformulação do modelo.

Dos tipos de simulação existentes¹⁴, terão especial interesse no nosso estudo a simulação estocástica, uma vez que as variáveis, das quais se procuram gerar amostras, seguem leis aleatórias, com enfoque específico no caso discreto.

¹⁴ De um lado a simulação associada a modelos determinísticos (cujos os dados são obtidos como certezas), do outro a simulação associada a modelos probabilísticos.

A simulação de valores para as variáveis aleatórias deu origem ao desenvolvimento dos métodos de Monte Carlo, cujo aperfeiçoamento computacional tem sido notável e com uma aplicabilidade em várias áreas (Estatística, Matemática, Economia, Finanças, Física, Biologia, ...). Estes tipos de métodos, basicamente, envolvem a geração de observações de determinada distribuição de probabilidades e o uso da amostra obtida para aproximar à função de interesse. Na simulação de Monte Carlo, cada variável de um modelo de avaliação é representada por uma função densidade de probabilidade (ou por um intervalo de valores possíveis) e não por um simples valor, como na avaliação determinística.

Com efeito, para haver simulação, necessitamos da função densidade de probabilidade, que descreve o comportamento do sistema matemático (real), a qual pode ser “determinada” com base em dados observados, experimentais ou teóricos.

No nosso caso, a anteceder a uma eficiente utilização do método de Monte Carlo na estimação de medidas de interesse (com a aproximação de Monte Carlo na resolução de problemas de integração no cálculo de um valor esperado), há necessidade de encontrar métodos de simulação de distribuições *a posteriori*. Os métodos dinâmicos de Monte Carlo revelam-se assim úteis, dada a necessidade de amostragem de valores para variáveis aleatórias de distribuições de interesse.

Com efeito, na presente secção são tecidas algumas considerações teóricas referentes aos métodos iterativos de Monte Carlo e, posteriormente, a métodos de Monte Carlo baseados em simulações de Cadeias de Markov, cuja distribuição estacionária de interesse é a distribuição *a posteriori*. Tal como referido, com o objectivo de gerar uma amostra, com as características necessárias à aplicação de Monte Carlo (requisitos a expor nos parágrafos subsequentes), são apresentados dois procedimentos possíveis (os algoritmos de *Metropolis–Hasting* e *Gibbs*), sublinhando-se, desde já, que muitos outros procedimentos poderiam ser explorados, pois a investigação neste campo, após o trabalho pioneiro de Gefand e Smith (1990), tem sido considerável.

Como tal, o objectivo será apresentar a base teórica indispensável à compreensão da implementação dos algoritmos supramencionados, sugerindo algumas referências bibliográficas para uma compreensão mais detalhada dos aspectos em discussão.

1.4.2. MÉTODO DE MONTE CARLO (MMC)

O MMC¹⁵, assente em procedimentos numéricos, começou a ser utilizado e estudado, com maior enfoque, no seguimento de trabalhos desenvolvidos por alguns autores, como Fermi, von Neumann e Ulam, em 1948.

¹⁵ Este método, embora já tenha sido usado anteriormente (por exemplo, existe um registo de um artigo escrito por Lord Kelvin, que utilizava já técnicas de Monte Carlo numa discussão das equações de

A utilização de MMC, sendo uma alternativa a procedimentos numéricos, desempenham um papel preponderante na estatística bayesiana, na medida em que o cálculo, via analítica/numérica, de quantidades *a posteriori* é por vezes muito complexo, em particular em situações que envolvem integração, dado que no caso em que o domínio de integração é de pequena dimensão, os procedimentos numéricos são passíveis de ser implementados, mas em cenários multidimensionais, problema típico em inferência bayesiana, o recurso a este método assume-se claramente vantajoso.

Este método, consistindo na substituição do estudo de um processo matemático por um modelo probabilístico que possa tratar problemas determinísticos por meio de amostras aleatórias (ou números pseudo-aleatórios¹⁶), tem por base simulações estocásticas, na medida em que reproduz valores de distribuições de probabilidades, isto é, valores simulados de quantidades aleatórias com distribuições discretas ou contínuas.

O que procuraremos fazer em seguida é uma descrição geral do MMC, enquadrando-o com o trabalho a desenvolver, nomeadamente no que respeita a aplicabilidade na inferência bayesiana, cujo problema fundamental é estimar uma média *a posteriori*. Para um estudo mais pormenorizado de técnicas de simulação estocástica e descrição de procedimentos básicos, pode ser consultada uma vasta bibliografia, como por exemplo Gamerman (1996), entre outros.

Começemos por apresentar o conceito que sustenta toda a apresentação – valor esperado. De um modo geral, à semelhança do definido em Koop (2004), se considerámos uma função de interesse, g , e uma variável aleatória discreta, $X = \{x_i : i = 1, \dots, n\}$, podemos definir o valor esperado de $g(X)$ como:

$$\mathbb{E}[g(X)] = \sum_{i=1}^n g(x_i) p(x_i)$$

onde, para cada $i = 1, \dots, n$, $p(x_i)$ representa uma fdp.

Esta noção revela-se fundamental, na medida em que está na base de métodos que visam obter uma aproximação de um integral, via Monte Carlo. No seguimento dos

Boltzmann), passou a ser conhecido com esta designação – Monte Carlo – a partir da década de 40, em plena Segunda Guerra Mundial. O nome "Monte Carlo" terá surgido no seguimento do projecto Manhattan, onde, na construção da bomba atómica, se considerou a possibilidade de utilizar o método, que envolvia a simulação directa de problemas de natureza probabilística. Sendo Monte Carlo um centro de casinos, apostas e jogos ("jogos de azar") e dada a similaridade existente entre este tipo de jogos e a simulação estatística, a bibliografia associa a este facto a origem do nome, em particular à famosa roleta de Monte Carlo, um dos dispositivos mecânicos mais simples de geração de números pseudo-aleatórios.

¹⁶ Embora tendo a aparência de serem gerados de uma forma aleatória (num espaço admissível), os valores são gerados, computacionalmente, de forma determinística.

pressupostos da inferência bayesiana, considere-se o problema inicial de obter uma aproximação do integral, num dado intervalo $[a, b]$.

$$\int_a^b f(x) dx \quad (1.14)$$

Vejamos dois métodos usados para a estimação do integral, MMC ordinário e MMC com amostragem de importância, de entre os que são sugeridos na literatura.

1.4.2.1. MMC ORDINÁRIO

No caso de ser possível escrever a função f como uma factorização na forma

$$f(x) = g(x) h(x)$$

onde $h(x)$ denota a fdp da variável aleatória, X , definida num intervalo $[a, b]$, com $x \in [a, b]$, o integral (1.14) pode ser escrito na forma:

$$\int_a^b g(x) h(x) dx = \mathbb{E}[g(X)] \quad (1.15)$$

Deste modo, o problema da avaliação do integral transforma-se num problema de estimação do valor esperado da variável aleatória $g(X)$, em relação à densidade $h(x)$.

Do ponto de vista da inferência bayesiana, sendo esse o nosso foco de interesse, assumindo que θ e x são conhecidos e $h(\theta|x)$ denota a função densidade *a posteriori*, a expressão (1.15), pode ser reescrita de uma outra forma, a qual vai servir de base a todas as observações subsequentes, isto é,

$$\int g(\theta) h(\theta|x) d\theta = \mathbb{E}[g(\theta)|x] \quad (1.16)$$

Os estimadores bayesianos são, geralmente, descritos em termos de valor esperado condicional de uma determinada função, g , integrável, pelo que a expressão (1.16) pode, em muitos casos, expressar uma quantidade *a posteriori*, como por exemplo as médias e as covariâncias *a posteriori* de componentes de θ , probabilidades *a posteriori* de subconjuntos do espaço de parâmetros e de densidades preditivas *a posteriori*. Além destas quantidades, existem outras de interesse no nosso caso específico, que podem ser

expressas por integrais do tipo (1.16) como é o caso das constantes de normalização das distribuições *a posteriori*, densidades marginais *a posteriori*, ou mesmo probabilidades *a posteriori* de modelos (Paulino *et al.*, 2003, p.286).

Neste seguimento, simulando-se uma amostra de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), $\theta_1, \dots, \theta_n \in \Theta$, da densidade *a posteriori*, $h(\theta|x)$, segundo o MMC, pode obter-se uma estimativa do integral escrito em (1.16) dada por uma média amostral:

$$\mathbb{E}[g(\theta)|x] = \frac{1}{n} \sum_{i=1}^n g(\theta_i) \quad (1.17)$$

Pelo que, conseguindo-se simulações da distribuição *a posteriori*, $h(\theta|x)$, a aplicação do MMC é imediata, uma vez que a utilização deste método, depende, essencialmente, da possibilidade de se gerarem pontos no domínio de integração, Θ , de acordo com a função densidade de probabilidade $h(\theta|x)$.

A garantia de que esta estimativa converge quase certamente (q.c.) para o valor pretendido, $\mathbb{E}[g(\theta)|x]$, é dada pela Lei Forte dos Grandes Números¹⁷. A este propósito, prova-se que (ver Paulino *et al.*, 2003) a precisão da aproximação supramencionada pode ser quantificada pelo erro padrão de Monte Carlo, isto é,

$$\varepsilon = \frac{\sqrt{\sum_{i=1}^n \left[g(\theta_i) - \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right]^2}}{\sqrt{n(n-1)}}$$

Sublinhe-se que, sendo um método baseado em simulação estocástica, resultado da simulação de uma amostra (aleatória) da distribuição base, o valor encontrado é uma aproximação, sendo a precisão da mesma melhorada com o aumento da dimensão da amostra, cujo único inconveniente reflecte-se na parte computacional. Porém, com o desenvolvimento dos métodos computacionais, os resultados da implementação dos MMC permitem obter valores quase exactos, sem grandes esforços do ponto de vista computacional.

Para uma análise mais detalhada sobre estes e outros aspectos, entre outras referências, poder-se-á citar-se Shonkwiler e Mendivil (2009), onde é exemplificada a aplicabilidade

¹⁷ Lei Forte dos Grandes Números:

Sejam X_1, \dots, X_n variáveis aleatórias i.i.d.. Para $i = 1, \dots, n$, $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2 < \infty$, defina-se a média $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Então, para cada $\varepsilon > 0$, \bar{X}_n converge q.c. para o seu valor esperado (esperança) μ , isto é, $P(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \varepsilon) = 1$.

deste método na simulação de várias distribuições de probabilidades, que, para além de deduções analíticas necessárias, são mostrados os códigos da implementação em MatLab e respectivos outputs, incluído a análise do erro.

O que se procurará fazer de seguida é uma referência mais particularizada, procurando estabelecer uma ponte entre os aspectos genéricos acima descritos e factos de interesse para a inferência bayesiana, decorrentes do exposto na Secção 1.2.

Para o efeito, considerando y o observado e θ como sendo o parâmetro do modelo que procura explicar y , recorde-se que:

$$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)}$$

onde, $p(\theta|y)$ denota a distribuição *a posteriori* de θ , depois de se observar y , $p(\theta)$ denota a distribuição *a priori* sobre θ e $p(y|\theta)$ é a função de verosimilhança. Justificou-se, ainda, que $p(\theta|y)$ depende na íntegra de $p(y|\theta)$ e $p(\theta)$, sendo proporcional ao produto das duas componentes, isto é:

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta)$$

O desafio com o qual nos deparamos, na econometria bayesiana, é estimar a média da distribuição *a posteriori* – média *a posteriori*:

$$\mathbb{E}[\theta|y] = \int \theta p(\theta|y) d\theta$$

Tal como refere Koop (2004), tal estimativa revela-se, por vezes, muito complexa.

Neste seguimento, se considerámos uma função de interesse, g , e o vector aleatório discreto, $\theta_1, \dots, \theta_n$ que procura explicar y , a aproximação de Monte Carlo dá-nos:

$$\mathbb{E}[g(\theta)|y] = \sum_{i=1}^n g(\theta_i) p(\theta_i)$$

No caso das variáveis aleatórias serem i.i.d., pela Lei Forte dos Grandes Números, temos a garantia que um estimador consistente para a média *a posteriori*, é dado por

$$\widehat{\mathbb{E}}[g(\theta)|y] = \frac{1}{n} \sum_{i=1}^n g(\theta_i)$$

Como forma de ilustrar a implementação computacional, podemos esquematizar as etapas que estão na base do algoritmo:

1° Passo	Gerar a amostra i.i.d., $\theta_1, \dots, \theta_n$, da distribuição de $p(\theta y)$;
2° Passo	Calcular $g(\theta_1), \dots, g(\theta_n)$;
3° Passo	Calcular a média amostral $\bar{g} = \frac{1}{n} \sum_{i=1}^n g(\theta_i)$ (estimador de Monte Carlo).

TABELA 1 – Esquema de computação do MMC Ordinário

1.4.2.2. MMC COM AMOSTRAGEM DE IMPORTÂNCIA

A estimativa de Monte Carlo anteriormente apresentada exige alguns requisitos, nomeadamente a necessidade de obter amostras i.i.d. a partir da distribuição *a posteriori*.

Apesar da diversidade de métodos de simulação, o problema é que, geralmente, não é possível obter amostras i.i.d. directamente da amostra de $h(\theta|x)$. A solução passa por simular de uma distribuição idêntica à distribuição *a posteriori*, sendo que alguns autores apresentam um método mais aplicável, conhecido na literatura temática como *importance sampling method*.

Paulino *et al.* (2003) descreve o MMC com amostragem de importância, considerando, para o efeito, uma função densidade $q(\theta)$ da qual é fácil simular valores, usualmente designada *função de importância*, verificando-se

$$h(\theta|x) \propto f(x|\theta) \cdot h(\theta).$$

Partindo do integral (1.16) pode escrever-se:

$$\begin{aligned} \int g(\theta) h(\theta|x) d\theta &= \frac{\int g(\theta) f(x|\theta) h(\theta) d\theta}{\int f(x|\theta) h(\theta) d\theta} \\ &= \frac{\int \left[g(\theta) \frac{f(x|\theta) h(\theta)}{q(\theta)} q(\theta) \right] d\theta}{\int \left[\frac{f(x|\theta) h(\theta)}{q(\theta)} q(\theta) \right] d\theta} \\ &= \frac{\int [g(\theta) w(\theta) q(\theta)] d\theta}{\int [w(\theta) q(\theta)] d\theta} \end{aligned}$$

Assim, aplicando o MMC, conseguida uma amostra $\theta_1, \dots, \theta_n$ de $q(\theta)$, obtém-se uma estimativa para a média *a posteriori* dada por:

$$\hat{\mathbb{E}}[g(\theta)|x] = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n g(\theta_i) w_i \quad (1.18)$$

onde, para cada $i = 1, \dots, n$,

$$w_i = \frac{f(x|\theta_i)h(\theta_i)}{q(\theta_i)}$$

A demonstração dos factos de interesse, face ao exposto, pode ser consultada em Geweke (1989). O autor refere que, mediante as hipóteses: (i) o suporte de $q(\theta)$ incluir o suporte de $h(\theta|x)$; (ii) o vector de variáveis aleatórias, $\theta_1, \dots, \theta_n$ ser uma amostra i.i.d. de $q(\theta)$; (iii) garantida a existência do integral (1.16) e este ser finito, então a estimativa (1.18) converge q.c. para o integral (1.16), isto é:

$$\frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n g(\theta_i) w_i \xrightarrow{nnnn} \int g(\theta) h(\theta|x) d\theta \quad q.c.$$

O autor apresenta, ainda, uma estimativa do erro padrão de Monte Carlo dada por:

$$\varepsilon = \frac{\sqrt{\sum_{i=1}^n \left[g(\theta_i) - \frac{1}{\sum_{j=1}^n w_j} \sum_{i=1}^n g(\theta_i) w_i \right]^2 w_i^2}}{\sum_{j=1}^n w_j}$$

Observe-se que, para a aplicação desta metodologia, basta exigir que $h(\theta|x)$ é conhecida e a precisão da estimativa (1.18) irá ser mais exacta à medida que $q(\theta)$ se aproxima da distribuição de $h(\theta|x)$. No caso (óptimo) em que $q(\theta)$ e $h(\theta|x)$ coincidem, a estimativa reduz-se ao caso directo da estima de Monte Carlo apresentada em (1.17).

Como forma de ilustrar a implementação computacional, uma solução aproximada, via Monte Carlo, pode ser obtida usando o seguinte algoritmo:

1º Passo	Gerar a amostra i.i.d., $\theta_1, \dots, \theta_n$, da distribuição de $h(\theta)$;
2º Passo	Calcular w_1, \dots, w_n , com $w_i = \frac{h(\theta_i x)}{q(\theta_i)}$;
3º Passo	Calcular a média amostral $\bar{g} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n g(\theta_i) w_i$ (estimador de Monte Carlo), com: <ul style="list-style-type: none"> ▪ $g(\theta) = \theta$ para o cálculo aproximado do valor médio da distribuição a posteriori; ▪ $g(\theta) = \theta^2$ para o cálculo aproximado da variância da distribuição a posteriori;

TABELA 2 – Esquema de computação do MMC com amostragem de importância

Para uma análise mais detalhada da aplicação de Monte Carlo na avaliação de probabilidades *a posteriori*, densidades *a posteriori* marginais, intervalos de credibilidade e quantidades associadas à distribuição preditiva *a posteriori*, nomeadamente das expressões que caracterizam a estimativa de Monte Carlo para cada caso, consulte-se Paulino *et al.* (2003).

1.4.3. MÉTODO DE MONTE CARLO VIA CADEIAS DE MARKOV (MCMC)

Na estimação, via MMC ordinário, do valor esperado de uma distribuição *a posteriori* de interesse, onde se obtém uma amostra da distribuição *a posteriori* num único passo, os valores são gerados de forma independente, não havendo preocupação com a convergência do algoritmo, dado que a mesma é automaticamente garantida desde que o tamanho da amostra seja suficientemente grande. Por isso, estes métodos são chamados não iterativos.

No entanto, em muitos problemas pode ser bastante difícil, ou mesmo impossível, encontrar uma densidade de importância que seja, simultaneamente, uma boa aproximação da distribuição *a posteriori* e fácil de ser amostrada (Ehlers, 2007). Daí se ter justificado a dificuldade da aplicação directa do método, nos casos em que obter amostras i.i.d., a partir da distribuição *a posteriori*, se revela algo complexo, deparando-nos com a necessidade de uma procura de métodos (não directos) que permitam uma “boa” simulação da distribuição de interesse. É neste seguimento que surgem alguns métodos dinâmicos (não iterativos), procurando dar resposta ao problema exposto, onde o objectivo é obter uma amostra da distribuição *a posteriori* e calcular estimativas amostrais de interesse dessa mesma distribuição.

“A ideia básica por detrás desses métodos é a de transformar o problema estático em consideração num problema de natureza dinâmica, construindo para o efeito um processo estocástico temporal, artificial, que seja fácil de simular, e que convirja, fracamente, para a distribuição original. Este processo temporal é, em geral, uma Cadeia de Markov homogénea cuja distribuição de equilíbrio é a distribuição que se pretende simular [que na inferência bayesiana é a densidade *a posteriori*]”. (Paulino *et al.*, 2003, p. 305)

Ou seja, este processo alternativo tem por base substituir a geração de realizações independentes de θ , de acordo com $h(\theta|x)$, por uma sucessão de realizações correlacionadas geradas a partir de uma Cadeia de Markov (com algumas características específicas), com um determinado espaço de estados, Θ , e uma distribuição estacionária específica.

Assim, com o uso de métodos MCMC é possível gerar uma cadeia de Markov homogénea e ergódica¹⁸ que admite uma distribuição de equilíbrio. Com efeito, após a convergência da cadeia para a distribuição de interesse, os valores gerados constituem uma amostra desta distribuição, podendo ser usados para “cálculos” via Monte Carlo.

O contributo dado por Metropolis em 1953, com o trabalho “*Equation of state calculations by fast computing machines*” e Hastings em 1970, com o trabalho “*Monte Carlo sampling methods using Markov chains*”, no desenvolvimento de um algoritmo para o efeito, acabaria por dar nome a um dos algoritmos mais referenciados na literatura sobre o tema – algoritmo de *Metropolis-Hastings*.

Posteriormente, o método de amostragem de *Gibbs*, cujas referências bibliográficas remetem para o trabalho desenvolvido por Geman e Geman – “*Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*” – em 1984, surge como um caso particular do algoritmo de *Metropolis-Hastings*. No entanto, a aplicabilidade deste método na Inferência Bayesiana, com a simulação de distribuições *a posteriori*, viria a ser feita por Gelfand e Smith (1990) em “*Sampling-based approaches to calculating marginal densities*”.

Em linhas gerais o objectivo é gerar uma cadeia de parâmetros, $\theta_1, \theta_2, \dots$, que converge¹⁹ para uma distribuição estacionária de interesse, a distribuição *a posteriori*,

¹⁸ Daqui em diante, sempre que nos referirmos à construção de uma Cadeia de Markov via métodos MCMC, subentende-se que a mesma é homogénea e ergódica, no seguimento do exposto na secção 1.3.

¹⁹ Recorde-se que a garantia de convergência foi justificada na secção 1.3. Assim, se as simulações iniciais (frequentemente chamado de período de *burn-in*) são ignoradas, a partir de certa ordem é garantida a convergência.

cadeia essa à qual podemos aplicar, posteriormente, o MMC directo para estimar o valor de medidas *a posteriori* expressas por meio de um integral.

Para uma melhor compreensão, far-se-á uma descrição dos aspectos teóricos que sustentam o algoritmo de *Metropolis-Hastings* e o método de amostragem de *Gibbs* (caso particular do primeiro), bem como um esquema da implementação computacional dos dois algoritmos. No entanto, reconhecendo que a análise feita é redutora, uma discussão mais geral sobre os métodos MCMC pode ser consultada, entre uma vasta bibliografia, em Gamerman e Lopes (2006) e para um enquadramento/descrição detalhada dos algoritmos veja-se Glynn e Asmussen (2007).

1.4.3.1. ALGORITMO DE *METROPOLIS-HASTINGS*

O algoritmo de *Metropolis-Hastings*, tal como outros métodos MCMC, visa obter uma amostra de uma distribuição *a posteriori*, $h(\theta|x)$, com a construção de uma cadeia de Markov, $\theta_1, \dots, \theta_n$, em Θ (espaço de estados), cuja distribuição de equilíbrio é a própria distribuição $h(\theta|x)$, sendo que, neste caso, um valor de uma distribuição auxiliar é gerado, podendo este ser aceite, ou não, com uma dada probabilidade, construindo-se assim uma cadeia, a qual converge para a distribuição de equilíbrio.

Como no nosso caso, a abordagem de interesse é para o caso discreto, fará sentido seguir a nomenclatura adoptada para o caso discreto, à semelhança da abordagem feita por Paulino *et al.* (2003) e Ehlers (2007).

Com efeito, a ideia geral do algoritmo é simular uma Cadeia de Markov, $(X_n)_{n \in \mathbb{N}_0}$, em Θ , que, após algum tempo de simulação, são amostra de uma distribuição aproximadamente igual a π .

Assim, dado um estado $\theta \in \Theta$, suponhamos que se deseja obter uma amostra de uma distribuição que atribui a cada estado uma probabilidade positiva $p(\theta) = \pi(\theta) = P(X = \theta)$ e, dados $\theta, \theta' \in \Theta$, $p(\theta'|\theta) = P(X_{k+1} = \theta' | X_k = \theta)$ ²⁰. Face ao exposto na Secção 1.3, sabemos que tal é possível se se conseguir encontrar uma Cadeia de Markov (com as características requeridas: homogénea e ergódica) que satisfaça as condições de reversibilidade (Definição (1.17)):

$$\pi(\theta) p(\theta'|\theta) = \pi(\theta') p(\theta|\theta'), \quad \forall \theta, \theta' \in \Theta \ (\theta \neq \theta')$$

²⁰ Observe-se que na secção 1.3 estas probabilidades foram denotadas por $p_i = \pi_i = P(X = i)$ e por $p_{ij} = P(X_{n+1} = j | X_n = i)$, sendo que neste caso procuraremos adaptar as notações ao caso de interesse.

Denotando por Q a matriz de transição da cadeia, onde cada elemento $q(\theta|\theta')$ representa a probabilidade de transição entre os elementos $\theta, \theta' \in \Theta$, considere-se que cada elemento $p(\theta|\theta')$ tem a forma:

$$p(\theta'|\theta) = q(\theta'|\theta) \alpha(\theta', \theta), \quad \forall \theta, \theta' \in \Theta \ (\theta \neq \theta')$$

com $0 \leq \alpha(\theta', \theta) \leq 1$, $\forall \theta, \theta' \in \Theta$.

Considerando que $X_k = \theta$, isto é, que num dado instante k a cadeia se encontra no estado θ , a simulação deste processo, em cada instante k , está assente na execução de duas etapas, as quais passamos a explicar:

- selecciona-se um estado θ' , cuja probabilidade de transição é dada pelo elemento $q(\theta'|\theta)$, usando a matriz Q ;
- o estado seleccionado é aceite, com probabilidade $\alpha(\theta', \theta)$, como o novo estado da cadeia no instante $k + 1$, ou rejeitado com probabilidade $1 - \alpha(\theta', \theta)$, sendo que nesse caso a cadeia permanece no estado θ .

Note-se que não é desejável escolher uma matriz de transição Q onde existe um número considerável de instantes em que $X_{k+1} = X_k$, ou seja, existe uma rejeição. Neste caso, a cadeia pode permanecer no mesmo estado por muitas iterações, algo que é passível de ser monitorizado, calculando a percentagem média de iterações para as quais novos valores são (ou não) aceites.

A questão que se coloca agora é qual o valor da probabilidade de aceitação que se deve considerar, sendo que a aceitação, ou não, resultará de uma comparação, para certos $\theta, \theta' \in \Theta$, do valor de α com um valor de uma função u .

Apresentam-se, de seguida, três exemplos de escolhas possíveis.

- I. O algoritmo sugerido por Metropolis em 1953 assume que, se a matriz Q for escolhida de modo a ser simétrica e se tomarmos

$$u(\theta', \theta) = 1 + \min\left(\frac{\pi(\theta)}{\pi(\theta')}, \frac{\pi(\theta')}{\pi(\theta)}\right)$$

então,

$$\alpha(\theta', \theta) = \min\left(1, \frac{\pi(\theta')}{\pi(\theta)}\right)$$

II. Uma generalização deste algoritmo é obtida se considerarmos uma matriz de transição, Q , arbitrária (não necessariamente simétrica) com

$$u(\theta', \theta) = 1 + \min \left(\frac{\pi(\theta) q(\theta'|\theta)}{\pi(\theta') q(\theta|\theta')}, \frac{\pi(\theta') q(\theta|\theta')}{\pi(\theta) q(\theta'|\theta)} \right)$$

tendo-se, nesse caso,

$$\alpha(\theta', \theta) = \min \left(1, \frac{\pi(\theta') q(\theta|\theta')}{\pi(\theta) q(\theta'|\theta)} \right)$$

III. O algoritmo sugerido por Hastings (1970), daí a designação de algoritmo de *Metropolis-Hastings* descrito, assume uma matriz de transição, Q , e $u(\theta', \theta)$ uma função simétrica²¹, escolhida tal que $0 \leq \alpha(\theta', \theta) \leq 1$, para quaisquer $\theta, \theta' \in \Theta$. Nestas condições o valor de $\alpha(\theta', \theta)$, vem dado por:

$$\alpha(\theta', \theta) = \frac{u(\theta', \theta)}{1 + \frac{\pi(\theta) q(\theta'|\theta)}{\pi(\theta') q(\theta|\theta')}}}$$

Note-se que, neste algoritmo, a distribuição de equilíbrio apenas é considerada no quociente $\frac{\pi(\theta)}{\pi(\theta')}$, pelo que basta conhecer π a menos de uma constante de proporcionalidade.

Acresce agora analisar se, nas condições enunciadas, o algoritmo gera uma amostra da distribuição π estacionária.

A unicidade e convergência para uma única distribuição de interesse são garantidas desde que a matriz Q seja a matriz de transição de uma cadeia de Markov ergódica. Pois nesse caso, tal como referido em Paulino *et al.* (2003), a matriz \mathcal{P} (formada pelos elementos $p(\theta'|\theta)$), representa uma matriz de transição de uma cadeia de Markov ergódica, desde que $\pi(\theta) > 0$, qualquer que seja $\theta \in \Theta$.

Para finalizar esta apresentação, podemos considerar um caso particular, tomando para o efeito a distribuição $u \sim U(0,1)$, para exemplificar, em termos práticos, o algoritmo de *Metropolis-Hastings*. Deste modo, o algoritmo pode ser especificado pelas seguintes etapas:

²¹ Um exemplo que é comum utilizar é a distribuição uniforme, $u \sim U(0,1)$, sendo esse que utilizaremos para exemplificar uma especificação do algoritmo.

1º Passo	Inicializar o ciclo iterativo em $k = 0$ e especificar um valor inicial $\theta^{(0)}$;
2º Passo	Gerar um novo valor θ' da distribuição $q(. \theta)$;
3º Passo	Calcular a probabilidade de aceitação $\alpha(\theta', \theta)$;
4º Passo	Gerar $u \sim U(0,1)$;
5º Passo	<p>Se $u \leq \alpha$,</p> <p style="padding-left: 40px;">aceitar o novo valor e tomar $\theta^{(k+1)} = \theta'$</p> <p>Caso contrário,</p> <p style="padding-left: 40px;">Rejeitar o novo valor e tomar $\theta^{(k+1)} = \theta$</p>
6º Passo	Incrementar o ciclo de k para $k+1$ e retomar o 2º passo

TABELA 3 – Esquema de computação do algoritmo de *Metropolis-Hastings* (MCMC)

Tendo em conta o esquema apresentado, observe-se que o algoritmo permite mudar uma componente da cadeia de cada vez. Porém, é importante referir que isso apenas facilita o processo, não comprometendo a validade do mesmo, uma vez que cada elemento é actualizado ciclicamente por uma ordem previamente fixada. Veja-se o exemplo 1.5 (Anexo A) ilustra a aplicação deste algoritmo.

1.4.3.2. MÉTODO DE AMOSTRAGEM DE GIBBS

Um algoritmo particular na classe de algoritmos apresentada é o algoritmo de *Gibbs*²².

“O método de amostragem Gibbs não é mais que um método de simulação de distribuições multivariadas de natureza bastante complexa, baseado na sua caracterização através das distribuições condicionais completas. O algoritmo de simulação proposto define uma cadeia de Markov que, sob condições muito gerais, tem como distribuição limite a distribuição conjunta que se pretende simular. Assim, se se simularem estados da cadeia por um longo período de tempo, de modo a garantir que o equilíbrio seja atingido, pode construir-se uma amostra da distribuição multivariada à custa dos estados da cadeia assim gerada”. (Turkman, s.d.)

²² O algoritmo de *Gibbs* foi introduzido por Josiah Willard Gibbs (físico, químico e matemático nascido nos EUA) em 1878 e interpretado mais tarde por Jaynes (1957) como uma técnica mais geral e aplicável na inferência. Porém, foi com o trabalho proposto por Geman e Geman (1984), enquanto método de amostragem para espaços discretos, generalizado, posteriormente, para espaços mais gerais por Gelfand e Smith (1990), que o algoritmo ganhou particular interesse entre os estatísticos, em particular bayesianos.

A particularidade deste método de amostragem é que não existe um mecanismo de aceitação/rejeição, isto é, não existe uma probabilidade de aceitação/rejeição, pelo que a cadeia irá sempre mover-se para um novo valor.

Conjugando algumas considerações feitas em Paulino *et al.* (2003) e Ehlers (2007), procurando adaptá-las ao nosso caso, se considerarmos π a distribuição de probabilidade conjunta e $\theta = (\theta_1, \dots, \theta_n)$, o algoritmo baseia-se no facto de que as transições entre estados são feitas de acordo com as distribuições condicionais completas

$$\pi(\theta_i | \theta_{-i}), \quad i = 1, \dots, n,$$

onde $\theta = (\theta_1, \dots, \theta_n)$ e θ_{-i} representa o vector θ sem a i -ésima componente, isto é, $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$. Com efeito, podendo cada um das componentes θ_i ser uni ou multidimensional, a distribuição condicional completa é a distribuição da i -ésima componente de θ condicionada em todas as outras componentes.

Deste modo, considere-se um algoritmo para gerar uma amostra de $\pi(\theta)$, onde, para cada $k = 1, \dots, n$, a transição de θ_k para θ_{k+1} é feita substituindo a j -ésima componente por um valor gerado pela distribuição $\pi(\theta_i | \theta_{-i})$, com $j = 1, \dots, n$. Note-se que este processo é um caso particular do algoritmo de *Metropolis-Hastings*, tomando a distribuição condicional completa como proposta e a probabilidade de aceitação $\alpha = 1$.

Assim, apesar de em certas condições, a geração de uma amostra directamente de $\pi(\theta)$ poder ser complexa, este método de amostragem é uma solução, podendo ser explicado no seguinte esquema de implementação, sendo que $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$ denota o vector inicial para θ .

1° Passo	Inicializar o ciclo iterativo da cadeia em $k = 0$;
2° Passo	Especificar valores iniciais para $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$
3° Passo	Obter um novo valor de $\theta^{(k+1)}$ a partir de $\theta^{(k)}$, procedendo iterativamente da seguinte forma: $\theta_1^{(k+1)} \sim \pi(\theta_1 \theta_2^{(k)}, \theta_3^{(k)}, \dots, \theta_n^{(k)})$ $\theta_2^{(k+1)} \sim \pi(\theta_2 \theta_1^{(k)}, \theta_3^{(k)}, \dots, \theta_n^{(k)})$ \vdots $\theta_n^{(k+1)} \sim \pi(\theta_n \theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_{n-1}^{(k)})$
4° Passo	Incrementar o ciclo de k para $k + 1$ e retomar o 2° passo

TABELA 4 – Esquema de computação do método de amostragem de *Gibbs* (MCMC)

Repare-se que, completo cada ciclo descrito no 3º passo, obtemos o novo vector $\theta^{(k+1)}$ e se se iterar n vezes este ciclo de geração de observações aleatórias de cada uma das distribuições condicionais (4º passo), produzimos a amostra $\theta^{(0)}, \dots, \theta^{(n)}$ desejada que é uma Cadeia de Markov com espaço de estados, Θ , que após convergência, é uma amostra de $\pi(\theta)$.

Para finalizar a nossa análise, veja-se o exemplo numérico (Exemplo 1.6.) de implementação deste método apresentado no Anexo A.

2. SÉRIES TEMPORAIS: FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA

Antecedendo à implementação dos procedimentos práticos a que nos propomos neste trabalho, é necessária uma reflexão sobre os aspectos principais inerentes ao estudo de séries temporais²³, procurando fundamentar os modelos econométricos a usar.

Reconhecendo a brevidade com que alguns assuntos são aqui tratados, à semelhança do feito no Capítulo 1, sempre que a exploração fica aquém do desejável, sugerem-se algumas referências bibliográficas para uma análise mais detalhada. Entre outras possíveis, sugere-se van Dijk *et al.* (2004) ou Johston e Dinardo (2001) para uma leitura mais detalhada sobre os métodos e modelos econométricos em geral, evidenciando uma boa integração dos conceitos da teoria económica com as técnicas econométricas e os testes utilizadas na estimação e validação dos diversos modelos. Já Kirchgässner e Wolters (2007) apresentam uma análise mais centrada no estudo de séries temporais. Nesta referência são apresentados alguns desenvolvimentos no que respeita a métodos utilizados no estudo de séries temporais, procurando estabelecer ligações entre os métodos e as aplicações, com a constante apresentação de exemplos empíricos.

2.1. INTRODUÇÃO: MODELAÇÃO EM ECONOMETRIA

Pelo facto de muitas variáveis económicas serem passíveis de quantificação, é possível, por meio de funções matemáticas (modelos matemáticos), caracterizar comportamentos, desenvolver estudos e avaliar hipóteses com vista a fazer previsões. A base da análise econométrica reside, assim, no confronto desses modelos com a realidade, os quais descrevem relações entre variáveis económicas. Porém, o desenvolvimento do trabalho econométrico carece da existência prévia de relações passivas de serem testadas econometricamente, sendo, a partir daqui, desenvolvida toda a nossa apresentação e

²³ Uma série temporal, ou sucessão cronológica, é um conjunto de observações tomadas em determinados instantes temporais com intervalos de tempo igualmente espaçados.

estudo, com ênfase nos modelos econométricos que visam analisar a referida relação entre variáveis.

De uma forma geral, qualquer modelo contém ‘quantidades’ que são observáveis e outras que não são directamente observáveis. Com efeito, a regressão, enquanto técnica estatística de modelação, assume a forma de um modelo probabilístico que incorpora um termo referente à componente determinística e outro que incorpora o erro aleatório, ou seja, a componente aleatória. Assim, se considerarmos a forma geral dos modelos a implementar, estes podem ser expressos pela seguinte equação:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t \quad (2.1)$$

onde y_t representa a variável *endógena* e o membro da direita a função a estimar, em que os x_{it} , com $i = 1, \dots, k$, representam as variáveis *exógenas*, os β_j , com $j = 0, \dots, k$, os coeficientes de regressão²⁴ e ε_t corresponde ao que se convencionou designar por erro do modelo. Esta última variável, ε_t , incorpora a parte do erro (resultante do facto de estarmos a considerar um processo estocástico) e os efeitos de outras variáveis não representadas no modelo, por serem ignoradas ou por não apresentarem uma relação estável com a variável endógena. No entanto, um modelo só é validado se a variável ε_t corresponder a um processo estocástico puro, ou seja, toda a informação relevante está contida na parte determinística²⁵.

Se consideramos uma amostra com dimensão N , o modelo identificado na equação (2.1), ignorando o termo constante, β_0 , corresponde, em termos matriciais, a escrever:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Ou, de um modo abreviado, vem:

$$Y = X\beta + \varepsilon$$

²⁴ O parâmetro β_0 é o ponto de intersecção com o eixo dos yy , usualmente designado como o *intercept*. Alguns autores, para simplificação da notação, não têm em conta este parâmetro, suprimindo esta constante da equação.

²⁵ Em Mendes (2010) é feita uma descrição pormenorizada dos passos a percorrer na construção, validação e utilização de um modelo, bem como os pressupostos a verificar na componente aleatória, para que o modelo seja adequado.

onde X denota matriz $(N \times k)$ de variáveis *exógenas*, Y o vector $(N \times 1)$ de variáveis *endógenas*, β o vector $(k \times 1)$ dos coeficientes de regressão e ε o vector $(N \times 1)$ respeitante aos erros do modelos.

Com efeito, as ‘quantidades’ que não são directamente observáveis são designadas por parâmetros, que denotamos por θ , os quais podem ser restringidos a um espaço de parâmetros Θ .

Atendendo à equação (2.1), é lícito escrever $\theta = (\beta, \varepsilon)$. Deste modo, qualquer valor particular para θ é definido em função dos dados observados e, como tal, o valor assumido para o parâmetro define uma estrutura particular para o modelo, o qual se espera que esteja o mais próximo possível da realidade.

Em suma, tal como referido por Lancaster (2004), poder-se-á dizer que a análise econométrica procura avaliar a consistência do modelo com as evidências e, posteriormente (assumindo a existência de um modelo consistente), analisar as estruturas definidas pelo modelo, com vista a fazer inferências para a tomada de decisões económicas, por exemplo.

Contudo, na prática, tende-se a inverter um pouco o processo, isto é, começa-se por assumir que o modelo é consistente com os dados, investiga-se posteriormente qual a estrutura mais provável à luz das evidências. Na análise bayesiana que iremos desenvolver, tal consiste em usar os dados para formar uma distribuição de probabilidades *a posteriori* para Θ .

Neste seguimento, para uma compreensão dos modelos econométricos usados no nosso estudo, inicialmente far-se-á uma referência aos modelos auto-regressivos (AR) e consequente extensão dos mesmos, denominada de modelo auto-regressivo vectorial (VAR – *Vector Autoregressive*), a qual será, posteriormente analisada do ponto de vista da inferência bayesiana, daí a designação BVAR, procurando responder ao problema da sobre-parametrização que afecta os modelos VAR. Dadas algumas limitações da aplicação directa dos modelos VAR (apenas a séries estacionárias), será feita uma análise aos mecanismos de correcção de erro (MCE), introduzindo-se assim os modelos VECM (*Vector Error Correction Model*), os quais apresentam como variáveis determinísticas os MCE.

Antecedendo à apresentação e análise dos modelos econométricos clássicos (abordagem clássica aos modelos auto-regressivos) e à sua extensão bayesiana, vejamos alguns conceitos base que se mostram fundamentais no estudo de séries temporais, como são exemplo os conceitos de estacionariedade e cointegração.

2.2. NOÇÕES GERAIS NO ESTUDO DE SÉRIES TEMPORAIS

Contrapondo com estudos precedentes, em que os modelos econométricos usados assumiam, à partida, que a origem dos dados provinham de processos onde a média e a variância eram constantes ao longo do tempo, não dependendo assim do período a que nos reportávamos, Granger e Newblond em “*Spurious regressions in econometrics*”, desenvolvido em 1974, alertaram para o facto de ao se estimar uma regressão, ainda que os valores dos respectivos testes *t* de *student* fossem significativos e que o coeficiente de determinação da regressão, R^2 , fosse elevado, correr-se-ia o risco de se obter o que ficou denominado de regressão espúria. Desta forma, apesar dos testes estatísticos do modelo de regressão se apresentarem significativos, não existe uma relação consistente entre as variáveis, pelo que os resultados não têm significado económico.

Assim, pensando numa regressão entre duas séries temporais, o problema exposto surge quando ambas as séries envolvidas exibem forte tendência, verificando-se um valor de R^2 alto, o qual se deve à presença da tendência e não ao verdadeiro relacionamento entre as duas séries. Daí, conhecidos os riscos das regressões espúrias, em estudos ulteriores passou a usar-se a abordagem definida por Box e Jenkins²⁶, procurando ‘estabilizar’ as sucessões (médias e variâncias constantes ao longo do tempo) antes da sua inclusão nos modelos, cuja análise é mais simples, pois obtém-se a estabilidade dos parâmetros estimados.

De um modo sucinto, como numa série temporal os dados são registados sequencialmente ao longo do tempo, espera-se que ela apresente correlação serial no tempo. Assim, a metodologia de Box-Jenkins, genericamente encontrada na literatura nos modelos ARIMA (*Auto Regressive Integrated Moving Average*), visa captar o comportamento da correlação serial ou autocorrelação entre os valores observados da série temporal, com vista a realizar previsões futuras.

Procurando esquematizar a estratégia utilizada para a construção de modelos, pela metodologia Box & Jenkins, veja o ciclo descrito na figura seguinte, no qual a estrutura do modelo é formada pelos próprios elementos da série. Para uma análise mais detalhada

²⁶ Durante a década de 60 os professores George E. P. Box e Gwilym M. Jenkins desenvolveram diversos trabalhos sobre séries temporais. Em 1970 publicaram o livro *Time Series Analysis, forecasting and control* (apesar da mais citada na literatura ser a versão lançada em 1976). Deixando como sugestão bibliográfica para uma análise mais detalhada sobre o tema, o grande contributo desse trabalho foi reunir as técnicas existentes numa metodologia para construir modelos que descrevessem com precisão (e de forma parcimoniosa) o processo gerador da série temporal, gerando previsões acuradas de valores futuros.

sobre os procedimentos esquematizados, veja-se o trabalho desenvolvido pelos autores em questão.

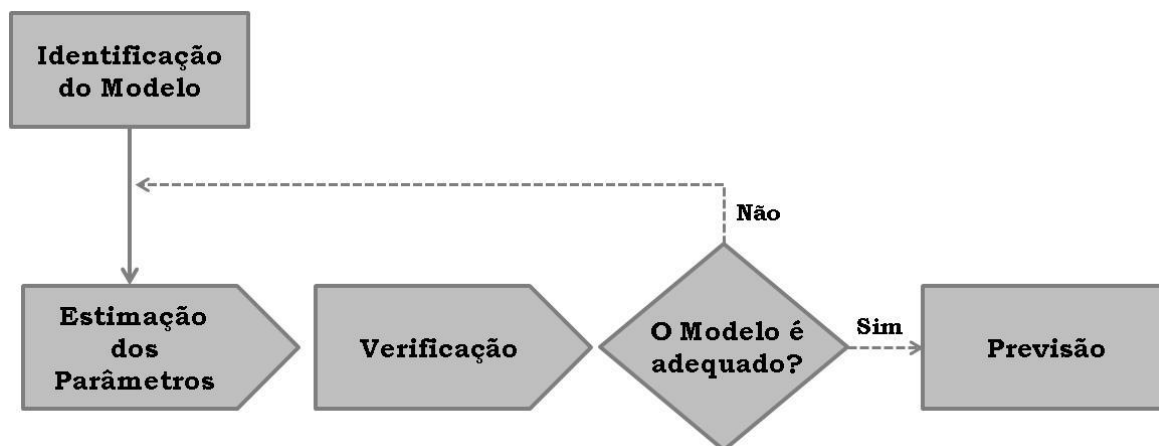


FIGURA 4 – Ciclo iterativo da construção de modelos de modelação (Box & Jenkins)

Em suma, das linhas precedentes ressalta, desde logo, um dos conceitos chave que está na base de uma das etapas iniciais a percorrer no estudo de séries temporais, o estudo da estacionariedade.

2.2.1. ESTACIONARIEDADE

No seguimento do exposto, podemos afirmar que nos modelos com séries temporais admitimos que as variáveis assumem valores gerados por meio de um processo estocástico. Sendo o objectivo desenvolver e utilizar modelos capazes de prever, interpretar e testar hipóteses relativamente a dados económicos e financeiros, para que as inferências estatísticas sejam válidas (pois desta propriedade dependem as propriedades dos estimadores), um dos pressupostos a ser validado é a estacionariedade das séries temporais utilizadas. Isto é, verificar se as mesmas seguem um processo estocástico com média, variância constantes ao longo do tempo e a covariâncias entre os valores desfasados da série depende apenas do valor do desfasamento (designado por lag^{27}), isto é, da “distância” temporal entre elas.

²⁷ Define-se o operador lag , L , como sendo $L(y_t) = y_{t-1}$, ou mais geralmente $L^s(y_t) = y_{t-s}$, com s inteiro. Por exemplo, o operador de diferenciação, Δ , pode escrever-se como $1 - L$, tendo-se:

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$$

Tipicamente operamos sobre uma série não com o operador lag , mas com o polinómio operador lag :

$$P(L)y_t = a_0 y_t + a_1 y_{t-1} + \dots + a_n y_{t-n} = \sum_{k=1}^n a_k y_{t-k}$$

De um modo formal, define-se série estacionária como:

DEFINIÇÃO 2.1

Uma série $\{y_t\}_{t \in T}$ diz-se fracamente estacionária²⁸ (ou estacionária em covariância) se:

(i) A média (valor esperado) em qualquer momento não depende desse exacto momento, sendo sempre constante, isto é

$$\mathbb{E}(y_t) = \mathbb{E}(y_{t-s}) = \mu, \quad \forall t, t-s \in T$$

(ii) A variância dos seus valores, seja qual o período a que se reporta, for constante e finita, isto é:

$$\mathbb{E}((y_t - \mu)^2) = \mathbb{E}((y_{t-s} - \mu)^2) = \sigma_y^2 < \infty, \quad \forall t, t-s \in T$$

(iii) A covariância entre as diferentes observações não depende do período em que são tomadas, sendo constantes, isto é:

$$\begin{aligned} \mathbb{E}((y_t - \mu)(y_{t-s} - \mu)) &= \mathbb{E}((y_{t-s} - \mu)(y_{t-j-s} - \mu)) = \\ &= \gamma(s), \quad \forall t, t-s, t-j-s \in T \end{aligned}$$

No quadro seguinte veja-se a representação gráfica de algumas variáveis estacionárias e outras não estacionárias.



FIGURA 5 – Exemplos de séries estacionárias e não estacionárias

O Gráfico III ilustra um caso muito particular de série estacionária designado de ruído branco²⁹, onde os elementos da série não são correlacionados e apresentam independência

²⁸ Quando provamos a estacionariedade das variáveis, referimo-nos, em regra geral, à estacionariedade fraca. Importa referir que a estacionariedade implica estacionariedade fraca, mas o recíproco não é necessariamente válido, apenas se verifica no caso da distribuição ser Gaussiana (Mendes, 2010).

²⁹ Diz-se que um processo ε_t é um ruído branco (*white noise*) se verifica (i) $\mathbb{E}(\varepsilon_t) = 0$; (ii) $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ (iii) $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$ com $i \neq j$.

serial (y é independente e identicamente distribuído com média nula e variância constante).

No caso particular de séries financeiras, sendo estas, na maior parte dos casos, não estacionárias, a questão que agora se coloca é como ‘estabilizar’ uma série não estacionária. A resposta não é imediata pelo facto de não existirem regras *a priori* para se poder estacionar uma série. Contudo, podemos referir dois processos bastante utilizados, na prática, para estacionar variáveis:

- Diferenças sucessivas: $\Delta y_t = y_t - y_{t-1}$; $\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}$
- Diferenças logaritmizadas: $\Delta \log y_t = \log y_t - \log y_{t-1}$.

A título de exemplo, na figura seguinte estão exemplificadas, sequencialmente, cada uma das transformações efectuadas, onde é visível a evolução na obtenção de um série estacionária.

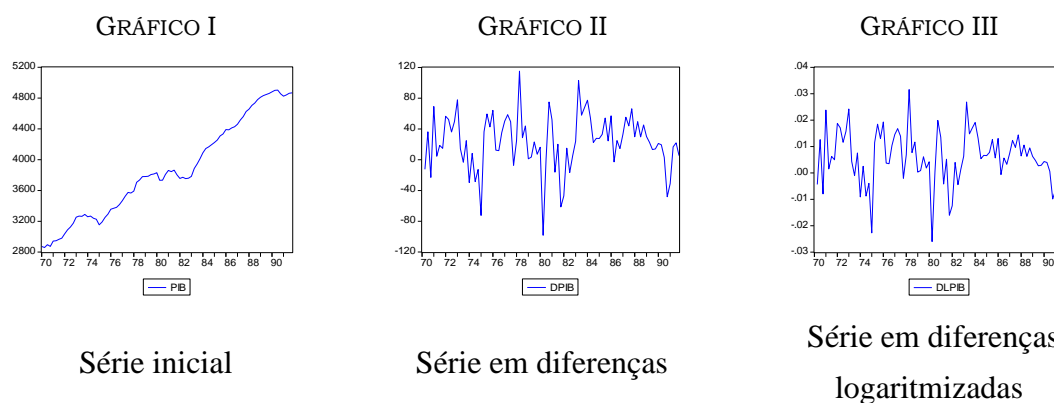


FIGURA 6 – Processos de transformação de uma série com vista à estacionariedade

No seguimento do primeiro processo apresentado para estacionar variáveis, surgem dois conceitos importantes no estudo de séries temporais: séries integradas e ordem de integração.

Uma série temporal é integrada de ordem d , escrevemos $y_t \sim I(d)$, se a série é não estacionária mas pode ser transformada numa série estacionária após d diferenças. Por exemplo, estatisticamente, uma série temporal será estacionária se a sua média, variância e covariância forem invariantes em relação ao tempo. Neste caso, a série será denotada por $I(0)$, significando que ela é integrada de ordem zero. Uma série que precisa ser diferenciada uma vez para atingir a estacionariedade é denotada por $I(1)$.

Em suma, a natureza da estacionariedade da série temporal tem influência directa, não só no seu comportamento, como também nas propriedades da mesma. Por um lado se numa série estacionária os choques se atenuam ao longo do tempo, numa série não

estacionária, os mesmos propagam-se. Além disso, os testes usuais das estimativas OLS³⁰ não fazem sentido numa regressão cujas variáveis não são estacionárias, pois são violados os pressupostos do modelo de regressão linear (ver Mendes, 2010).

2.2.2. RAIZ UNITÁRIA

Tal como se depreende dos exemplos apresentados na secção anterior, pode prever-se da análise gráfica se uma série é, ou não, estacionária. No entanto, pelo peso relativo de tal análise, dada a incerteza a ele inerente, devem realizar-se testes para aferirmos sobre a estacionaridade de séries temporais.

Assim, o propósito que nos propomos analisar de seguida é como detectar a (não) estacionaridade de uma série temporal, bem como os tipos de (não) estacionaridade existentes.

Podendo uma série temporal apresentar uma tendência determinística ou estocástica³¹, se pensarmos em termos de regressão usual

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (2.2)$$

a tendência estocástica pode surgir exclusivamente na variável exógena, x_t , nos erros, ε_t , ou então nas duas componentes, cumulativamente.

Com efeito, a questão chave é verificar se a tendência de uma série é determinística ou estocástica.

Para melhor analisar e discutirmos os aspectos seguintes, considere-se o seguinte modelo:

$$z_t = \rho z_{t-1} + \beta_0 + \beta_1 t + u_t \quad (2.3)$$

Temos quatro casos a considerar:

- Se $\rho = 0$, obtém-se um processo estacionário em tendência puro, isto é, um processo estacionário em torno de uma tendência linear, pelo que existe tendência determinística;

³⁰ O Método dos Mínimos Quadrados, ou OLS (*Ordinary Least Squares*), enquanto técnica de optimização matemática é a forma de estimação mais amplamente utilizada na econometria. Sintetizando, este método procura encontrar o melhor ajustamento para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados, designadas de resíduos.

³¹ Uma série apresenta uma tendência determinística se esta apenas varia em função do tempo, isto é, $y_t = \beta_0 + \beta_1 t + \varepsilon_t$, ao passo que se esta tiver a forma $z_t = z_{t-1} + \alpha + \varepsilon_t$, então dizemos que apresenta uma tendência estocástica.

- Se $\rho = 1$ e $\beta_1 = 0$, existe tendência estocástica, obtendo-se um processo estacionário em primeiras diferenças puro, o qual se designa por passeio aleatório. No caso de $\beta_0 \neq 0$ temos um passeio aleatório com *drift*³² (variações suaves); mas se $\beta_0 = 0$, obtém-se um passeio aleatório sem *drift* ($z_t = z_{t-1} + u_t$, exemplo clássico de série não estacionária);
- Se $\rho = 1$ e $\beta_{i=0,1} \neq 0$, obtemos um passeio aleatório com *drift* e tendência determinística;
- Se $0 < \rho < 1$, existe correlação serial em z (processo autorregressivo), mas não existe tendência estocástica.

Deste modo, estudar uma série quanto à estacionaridade, é verificar se $\rho = 1$, isto é, testar a existência de raiz unitária (designação usual). Assim, ao efectuar a regressão de z_t em z_{t-1} , a ideia associada aos testes de raiz unitária é verificar se o coeficiente estimado, ρ , é estatisticamente igual a 1.

TESTE DE RAIZ UNITÁRIA		
$\underbrace{H_0: \rho = 1}_{(não\ estacionária)}$	vs	$\underbrace{H_1: \rho < 1}_{(estacionária)}$

Com efeito, caso se tenha $\rho = 1$, dizemos que existe uma raiz unitária, ou seja, o processo é não estacionário.

De entre possíveis variantes dos testes de raiz unitária existentes e referidas na literatura, apresentamos aqui apenas duas variantes: os testes de raiz unitária de Dickey-Fuller (DF) e de Dickey-Fuller Aumentado (ADF), dado que serão os que iremos usar no nosso estudo e é os mais citados na bibliografia³³.

Nos testes de DF e ADF, para se poder inferir sobre a rejeição, ou não, de H_0 , não podemos utilizar directamente o valor da estatística t , dado que a distribuição de DF não é assimétrica, havendo necessidade de determinar os valores críticos próprios³⁴. Com efeito, caso o valor do teste seja superior aos valores críticos determinados para os níveis de confiança de 1%, 5% e 10%, não rejeitamos a hipótese nula, existindo uma raiz unitária, isto é, a série é não estacionária.

³² Uma das características destes processos é a persistência dos choques aleatórios (processos com memória infinita). O efeito de cada termo de erro não se dissipa ao longo do tempo, o processo guarda a informação de todos os choques sofridos até ao período corrente.

³³ Para mais detalhes sobre outros testes (por exemplo o teste de Phillips-Perron, comumente referido), consultar Kirchgässner e Wolters (2007).

³⁴ Importa referir que os valores críticos determinados nos testes DF e ADF não são os mesmos.

O modelo definido na equação (2.3), subtraindo z_{t-1} em ambos os membros, é dado por:

$$z_t - z_{t-1} = (\rho - 1)z_{t-1} + \beta_0 + \beta_1 t + u_t \quad (2.4)$$

Com efeito, temos:

TESTE DE RAIZ UNITÁRIA DE DICKEY-FULLER		
$H_0: \alpha = \rho - 1 = 0$ (<i>não estacionária</i>)	vs	$H_1: \rho < 0$ (<i>estacionária</i>)

Se o teste DF é utilizado no caso dos erros, u_t , não terem correlação serial, o teste de ADF é o utilizado quando os erros apresentam correlação serial.

Na prática, o teste ADF distingue-se do teste DF pelo facto de introduzir na equação desfasamentos (*lags*). Deste modo, nos testes de ADF é considerado um ‘novo’ modelo que se obtém de (2.4) adicionado valores desfasados, Δz , das variáveis independentes (neste caso dois *lags*), vindo:

$$\Delta z_t = z_t - z_{t-1} = (\rho - 1)z_{t-1} + \beta_0 + \beta_1 t + \beta_2 \Delta z_{t-1} + \beta_3 \Delta z_{t-2} + u_t \quad (2.5)$$

A questão prática que se coloca, neste caso, é qual o número de *lags* a adicionar. Schwert (1989) sugere uma regra empírica para a determinação do número de *lags*, k , dada por:

$$k = \left\lceil 12 \left(\frac{N}{100} \right)^{1/4} \right\rceil$$

onde N denota o tamanho da série e $[x]$ a parte inteira de x .

Uma observação final está relacionada com a aplicabilidade dos testes de DF e ADF, os quais devem ser aplicados levando em consideração cada uma das possibilidades, ou seja, considerando-se as seguintes hipóteses, com vista à obtenção de resultados mais conclusivos.

$H_0: z_t$ é um passeio aleatório	vs	$H_1: z_t$ é estacionário com média nula
$H_0: z_t$ é um passeio aleatório com <i>drift</i>	vs	$H_1: z_t$ é estacionário com média $\frac{\beta_0}{1-\rho}$
$H_0: z_t$ é um passeio aleatório com <i>drift</i> em torno de tendência determinística	vs	$H_1: z_t$ é estacionário em torno de tendência determinística

2.2.3. REGRESSÃO ESPÚRIA

Os testes de raiz unitária podem igualmente ser aplicados à componente residual, sendo que parte das considerações tecidas na secção anterior podem ser facilmente adaptadas para a variável que incorpora os erros do modelo.

Assim, o teste ADF cujas hipóteses nula e alternativa são as descritas anteriormente, pode ser igualmente usado para aferir sobre a estacionariedade da série dos resíduos, e_t . Com efeito, temos:

$$\Delta e_t = (\rho - 1)e_{t-1} + \beta_0 + \beta_2 \Delta e_{t-1}$$

O estudo da estacionariedade dos resíduos conduz a um outro conceito fundamental a que já nos referimos na introdução deste capítulo – regressão espúria.

Com efeito, considerando o modelo geral de regressão descritas em (2.2), no caso dos processos x e y serem, à partida, estacionários, o modelo de regressão clássico pode ser aplicado de imediato. Caso contrário, identificada a ordem de integração de cada um dos processos, sendo estes integrados da mesma ordem³⁵, far-se-á o estudo à componente residual. Caso esta seja estacionária, então dizemos que as séries são cointegradas. Caso contrário, isto é, caso contenha uma tendência estocástica, estamos perante uma situação de regressão espúria, sendo necessário proceder a mecanismos que permitam corrigir o erro para que o modelo estimado seja consistente.

Uma das estratégias usadas para ‘remediar’ uma regressão espúria é a estratégia de Newbold-Granger que, de um modo simplista, ao considerar as diferenças nas várias componentes, isto é:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \Delta \varepsilon_t \quad (2.6)$$

permite obter um modelo de regressão linear (a variável independente e os resíduos não contêm tendência estocástica).

³⁵ Caso os processos sejam integrados de ordem diferente, o modelo de regressão obtém resultados sem significado.

2.2.4. COINTEGRAÇÃO

Na secção anterior fizemos já referência ao conceito de cointegração, como propriedade estatística que garante a existência de uma relação de equilíbrio a longo prazo, não espúria, entre séries temporais não estacionárias.

Matematicamente, pensando na equação (2.2), com x e y integrados da mesma ordem e componente residual estacionária, esta pode ser escrita como $y_t - \beta_1 x_t = \beta_0 + \varepsilon_t$. Assumindo que a componente residual não tem tendência estocástica, então $y_t - \beta_1 x_t$ é um processo estacionário, ou seja, as variáveis x e y são cointegradas.

Assim, partindo de duas séries (ambas com tendência estocástica, mas integradas da mesma ordem), obtivemos uma combinação linear que resulta num processo estacionário, pelo que as variáveis são cointegradas. Neste caso, do ponto de vista gráfico, embora as séries se passem de forma estocástica, estas ficam próximas uma da outra, havendo convergência comum a longo prazo. Precisamente, o ajustamento feito através do mecanismo de correcção de erro (MCE) que faz com que as variáveis acabem por permanecer próximas, quando estas vagueiam aleatoriamente, permite identificar o vector de cointegração. Na figura seguinte, temos duas representações gráficas, onde em cada uma delas está representado um par de séries temporais. No primeiro caso, verificamos que, a longo prazo, as séries seguem caminhos diferentes, ao passo que no segundo caso, apesar das séries se passem de forma estocástica, existe uma convergência comum a longo prazo.

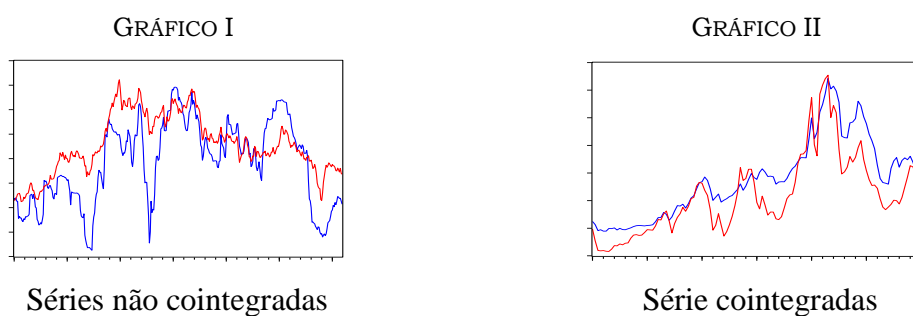


FIGURA 7 – Exemplo de séries cointegradas e não cointegradas

Estando interessados em estudar o comportamento a longo prazo, existe a necessidade de reescrever a equação (2.2) incluindo, não só as primeiras diferenças (nesse caso, a longo prazo, a equação anular-se-ia), mas também as variáveis em níveis, isto é:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + \varepsilon_t \quad (2.7)$$

onde o termo cointegrante, $y_{t-1} - \gamma x_{t-1}$, denota a correcção do erro³⁶ e o seu coeficiente, β_2 , traduz a velocidade de ajustamento (mudança de y para fazer face ao desequilíbrio).

Com isto temos apresentado o conceito de cointegração entre duas séries temporais, conceito que expressa a existência de um equilíbrio de longo prazo para o qual o modelo estimado converge no tempo.

De um ponto de vista mais formal, o conceito de cointegração foi desenvolvido, em 1987, por Engle e Granger no trabalho “*Cointegration and error correction: representation, estimation and testing*”, ao estabelecerem uma equivalência entre este conceito e os modelos de MCE – Teorema de Representação de Granger (como é conhecido na literatura).

No entanto, apesar das potencialidades reconhecidas na abordagem de Engle-Granger, existem algumas limitações. Para além da obtenção de diferentes resultados na estimação da regressão de cointegração para cada hipótese ensaiada, é possível que nos deparemos com a existência de múltiplos vectores de cointegração, obtendo-se, uma combinação linear dos diferentes vectores de cointegração, quando se utilizam mais de duas variáveis no modelo – modelo multivariado.

No sentido de ultrapassar tal problema (existência de mais do que um vector cointegrante), Johansen e Juselius, nos trabalhos desenvolvidos em 1989 e 1990, respectivamente, apresentam uma técnica baseada no método da máxima verossimilhança, a qual permite não só a estimação de todos os vectores cointegrantes, para um determinado conjunto de variáveis, como também testar restrições sobre os coeficientes dos vectores cointegrantes.

Pelo facto da técnica proposta estar baseada na representação do MCE de um modelo de vectores auto-regressivos, discutidos na secção seguinte, far-se-á uma análise matemática mais detalhada desta metodologia, nas páginas subsequentes deste trabalho. Quanto às etapas de implementação de cada uma das metodologias, estas serão observadas, pertinentemente, na parte prática desta dissertação.

No entanto, podemos desde já confrontar as duas propostas de abordagem: o Método de Engle-Granger (análise bivariada) e o Método de Johansen, desenvolvido por Johansen e Juselius (análise multivariada), pondo em relevo algumas particulares específicas de cada um métodos.

³⁶ Observe-se que, sendo x e y cointegradas, o termo $y_{t-1} - \gamma x_{t-1}$ é estacionário, pelo que pode ser estimado usando OLS.

Assim, em linhas gerais, temos:

MÉTODO DE ENGEL E GRANGER	MÉTODO DE JOHANSEN
<ul style="list-style-type: none"> ▪ Aplicável a modelos bicariados; ▪ Abordagem em dois passos (OLS), com base nos resíduos estimados; ▪ Ao ser utilizados para dados bivariados, assume, <i>a priori</i>, que existe um só vector de cointegração; ▪ Os ‘resultados da cointegração’ podem mudar em função da variável dependente seleccionada e das hipóteses consideradas na estimação da regressão de cointegração. 	<ul style="list-style-type: none"> ▪ Aplicável a modelos multivariados; ▪ Abordagem baseada em vectores autorregressivos (VAR); ▪ Testa a existência de múltiplos vectores cointegrantes, usando o Teste do Traço e o Teste do Valor Máximo; ▪ Técnica fundamentada nos testes de máxima verosimilhança, requerendo um volume significativo de dados.

2.2.5. CRITÉRIOS DE COMPARAÇÃO / SELECÇÃO DE MODELOS

Para finalizar esta abordagem geral a conceitos e ideias chave no estudo de séries temporais, vejamos alguns critérios usados numa das etapas subjacentes à modelação.

Uma das questões mais importantes no que respeita aos modelos auto-regressivos é precisamente o da escolha da sua ordem. Não existindo uma resposta, unanimemente aceite, é necessário recorrer a testes/critérios que nos permitam fazer opções com vista à escolha, o mais adequada possível, da ordem do modelo, dada a extrema importância de uma escolha apropriada.

Nesta linha, diversos critérios /estatísticas são apresentados na literatura para selecção de modelos, tendo estes como objectivo equilibrar o risco da escolha, por um lado de uma ordem menor que a real, provocando tal facto inconsistência na estimação dos parâmetros, por outro, de uma ordem superior, conduzindo ao incremento da variância desses estimadores. O equilíbrio em causa é feito através da atribuição de um custo/penalização pela exclusão/introdução de variáveis adicionais.

Na prática, os critérios baseados no máximo da função de verosimilhança são os mais utilizados, com maior ênfase para o Teste da Razão de Verosimilhança (LR), Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiana (BIC). Se o LR é apropriado para testar dois modelos, desde que um dos modelos seja um caso especial do

outro, qualquer um dos dois critérios comparam o ajuste dentro da amostra, medido pela variância dos resíduos, contra o número de parâmetros estimados.

Analiseemos, ainda que de um modo muito superficial, cada um dos três critérios acima referidos.

2.2.5.1. TESTE DA RAZÃO DE VEROSIMILHANÇA

No que respeita a modelos de regressão linear, existe um claro interesse em saber o valor da Soma dos Quadrados dos Resíduos (SQR), sendo que um valor alto da SQR sugere que a variável independente é importante, caso contrário, a variável independente não é útil na previsão da variável dependente.

Na regressão logística a ideia é a mesma: comparar os valores observados da variável dependente com os valores obtidos dos modelos (previsões) com e sem uma dada variável. A comparação dos valores observados com os valores preditos é baseada no logaritmo da verosimilhança, requerendo o LR a estimação do modelo restrito (sem inclusão da variável no modelo) e não restrito (com a inclusão da variável no modelo).

Com efeito, denotando por θ_s e por θ_c os vectores de parâmetros sem e com inclusão da variável no modelo, respectivamente, podemos calcular o valor maximizado da função verosimilhança sem e com a variável, denotado por $L(\theta_s)$ e $L(\theta_c)$, respectivamente.

De um modo mais formal, estando o teste LR baseado no logaritmo da razão entre as duas verosimilhanças, isto é, na diferença entre o $\ln(L(\theta_s))$ e $\ln(L(\theta_c))$, temos:

$$LR = -2[\ln(L(\theta_s)) - \ln(L(\theta_c))]$$

Assim, para testar a existência de relação, caso o valor da estatística seja maior que o valor crítico ao nível de significância desejado, rejeitamos a hipótese nula de exclusão no modelo.

2.2.5.2. CRITÉRIO DE INFORMAÇÃO DE AKAIKE

O Critério de informação Akaike foi desenvolvido pela Hirotugu Akaike, tendo sido publicado, pela primeira vez, em 1974 no trabalho "*A new look at the statistical model identification*".

Esta estatística, permite fazer uma selecção do(s) modelo(s) que melhor se ajusta(m), pois avalia a qualidade de ajuste do modelo estimado. Contrariamente a outros testes já referidos, esta estatística não fornece um teste de um modelo no sentido usual de testar

uma hipótese nula que traduza a adequabilidade do modelo aos dados num sentido absoluto.

O valor da estatística referente ao AIC incorpora duas componentes, uma que valoriza a precisão do ajuste e uma outra que penaliza, por meio de uma função crescente, os modelos com um maior número de parâmetros. Deste modo, a variação nos valores da estatística de AIC, relativa a cada um dos modelos em análise, deve-se ao logaritmo das razões de verossimilhança entre os modelos, penalizando-se os modelos pelo número de parâmetros.

Em sentido lato, sendo q o número de parâmetros do modelo e L o valor maximizado da função de verossimilhança, o valor de AIC é dado por:

$$\text{AIC} = -2 \ln(L) + 2q$$

Em suma, o AIC representa a distância relativa do modelo proposto ao modelo verdadeiro, ou de uma outra forma, a discrepância no ajuste do modelo proposto aos dados. Com efeito, o melhor modelo é aquele que apresenta um valor de AIC mais baixo.

2.2.5.3. CRITÉRIO DE INFORMAÇÃO BAYESIANA

O Critério de Informação Bayesiana, por vezes também referido como o Critério de Informação de Schwarz (SIC ou SC) por ter sido proposto por Gideon E. Schwarz no trabalho “*Estimating the Dimension of a Model*”, em 1978, é um dos mais utilizados na comparação de modelos. Este critério baseia-se, em parte, na função de verossimilhança, estando intimamente relacionado com AIC, mas incorporando um formalismo bayesiano. Se ambos os critérios permitem a comparação entre modelos, penalizando os modelos com maior número de parâmetros, no modelo bayesiano de Schwarz, essa penalização é mais rigorosa.

Basicamente, a ideia associada é partir-se de um modelo de regressão com vários *lags* e, gradualmente, ir reduzindo o número de *lags* até que se encontra aquele valor (número de *lags*) que minimize o valor do BIC.

De um modo mais formal, o valor da estatística associada a BIC é descrito por:

$$\text{BIC} = -2 \ln L + q \ln n$$

sendo q o número de parâmetros do modelo, n a dimensão da amostra e L o valor maximizado da função de verossimilhança.

Se tivermos em conta a perspectiva bayesiana, o termo referente à função verosimilhança (identificado na expressão anterior), tendo em conta o exposto na secção 1.2., ganha uma forma característica descrita na expressão seguinte:

$$\text{BIC} = -2 \frac{\sum_{i=1}^I \ln(f(y|\theta_i))}{I} + q \ln(n)$$

onde:

- $f(y|\theta_i)$ é a função de verosimilhança
- θ_i é o vector de parâmetros θ na i -ésima iteração (num total de I iterações)
- q é o numero de parâmetros do modelo
- n é a dimensão da amostra

Observe-se que o primeiro termo do BIC avalia o ajuste do modelo e o segundo termo corresponde à penalização de acordo com o número de parâmetros. Com efeito, é possível inferir que quanto menor for o valor do BIC, melhor será o modelo, tal como no AIC.

2.2.5.4. COMPARAÇÃO DOS CRITÉRIOS

Face às considerações feitas em cada caso, podemos concluir que os três critérios apresentados, apesar de conceitualmente diferentes no que respeita à forma de avaliar os modelos, utilizam o mesmo critério estatístico (maximização da função de verosimilhança) como medida do ajustamento, definindo valores críticos diferentes. Esta é a diferença fundamental entre os três processos.

Com o LR, considera-se por hipótese que o modelo mais simples é o de melhor ajuste, até que, dado um nível de significância α , se observem diferenças estatísticas para um modelo mais completo.

Por seu lado, utilizando-se o AIC admite-se que, do conjunto de modelos avaliados, nenhum modelo é considerado “modelo verdadeiro”, isto é, o que realmente descreve a relação entre a variável dependente e as variáveis explicativas, pelo que se tenta escolher o modelo que minimize as divergências. Já no BIC, está implícito que existe o modelo que descreve a relação entre as variáveis envolvidas (um “modelo verdadeiro”) e o critério tenta maximizar a probabilidade de escolha do verdadeiro modelo. Sublinhe-se finalmente, quando confrontamos os dois critérios, o modelo seleccionado usando a estatística BIC tende a ser mais parcimonioso comparativamente ao seleccionado pelo AIC.

2.3. MODELOS ECONOMETRICOS: VAR e VECM

Como se depreende de algumas observações feitas anteriormente, uma das áreas de pesquisa de séries temporais, que tem merecido especial interesse é a de modelos multiequacionais. O conceito de *Vector Auto-regressivo* generaliza os métodos univariados (adiante descritos como modelos AR) por considerar sistemas de equações simultâneas. Daí, ser lícito afirmar que os modelos VAR exploram as vantagens da análise multivariada, ultrapassando as limitações de modelos univariados. Apesar de ateóricos, os modelos em análise fornecem informações de causalidades entre as variáveis, decomposição da variância dos resíduos e a função de impulso resposta³⁷, o que assegura formas de entender as inter-relações entre as variáveis económicas. Além disso, o modelo capta o valor desfasado das variáveis, utilizando conjuntamente a inter-relação entre as variáveis com os consequentes desfasamentos.

Segundo a literatura sobre o tema, os modelos VAR foram introduzidos por Sims em 1980 com o trabalho *Macroeconomics and Reality*. Nesse artigo, o autor faz críticas aos modelos auto-regressivos existentes até então, propondo o VAR como alternativa. Segundo Sims (1986), nos modelos ditos tradicionais, por vezes, perdiam-se informações relevantes ao serem impostas restrições (desnecessárias) aos modelos econométricos, algo que é corrigido nos modelos VAR, pois as condições impostas *a priori* são poucas.

Desde então, estes modelos, enquanto modelos de séries temporais, têm sido amplamente utilizados na descrição das características estocásticas de séries económicas e na realização de previsões.

De um modo sucinto, esta metodologia utiliza algumas premissas sobre a série a ser analisada, conseguindo captar o efeito auto-regressivo de uma variável sobre si mesma e, também, o efeito de outras séries sobre a mesma variável. Além disso, possibilita a estimação de diversas variáveis, simultaneamente, evitando, problemas de identificação dos parâmetros em modelos multiequacionais. Estes factos, entre outros, sustentam a tese de que este tipo de modelos se revelou inovador, na medida em que constitui um importante instrumento de análise e estimação de equações com duas ou mais variáveis.

Antecedendo à apresentação dos modelos VAR, descreva-se primeiro o processo auto-regressivo – AR.

³⁷ Esta função traduz o comportamento de uma variável quando uma outra variável do sistema, ou ela própria, sofre um choque num dado instante que se irá reflectir nos instantes futuros. Tal permite prever o que acontece a cada variável, mediante um choque em qualquer outra variável do modelo.

No caso univariado, uma série temporal é modelada apenas em função dos seus valores passados e algumas perturbações, podendo esta ser escrita na forma

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-k}) + u_t$$

Porém, para que a expressão se torne operacional, é necessário a forma funcional da função f , o número de *lags* e uma estrutura para o termo perturbação, u (Johston e Dinardo, 2001). Com efeito, se considerarmos uma função linear com apenas um *lag* e uma perturbação do tipo ruído branco, obtemos um processo auto-regressivo de primeira ordem, AR(1), definido por:

$$y_t = a_0 + a_1 y_{t-1} + u_t \quad (2.8)$$

onde a_0 e a_1 são os parâmetros a estimar.

O processo definido em (2.8) pode ser generalizado se consideramos para o efeito um processo de ordem p , AR(p), o qual pode ser escrito na forma:

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + u_t \quad (2.9)$$

Embora não seja objecto de estudo no nosso trabalho, observe-se que quando a perturbação é um processo ruído branco o processo definido diz-se um *processo* AR(p) *puro*. Mas é possível assumir uma estrutura mais complexa para o termo perturbação, sendo que, a especificação usualmente assumida é um processo de médias móveis (MA), vindo o termo u_t , com ordem q , escrito na forma

$$u_t = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \dots + \delta_q \varepsilon_{t-q}$$

onde ε_t é um ruído branco.

Tal como referido, os processos derivados dos factos citados, ARMA/ARIMA, não serão explorados no nosso estudo. No entanto, para uma análise detalhada de processos univariados, veja-se Kirchgässner e Wolters (2007) ou Johston e Dinardo (2001).

Considerando agora uma extensão do modelo AR definido em (2.9), podemos definir os modelos VAR. Para o efeito basta considerar um vector coluna com k variáveis distintas, $\mathbf{y}_t = [y_{1t}, \dots, y_{kt}]'$, e modelá-lo em função dos seus valores passados, resultando assim uma autorregressão vectorial. Assim, o processo VAR(p) vem definido por:

$$\mathbf{y}_t = \boldsymbol{\varphi} + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2.10)$$

onde:

- \mathbf{y}_t é um vector coluna com k variáveis;
- $\mathbf{A}_i, i = 1, \dots, p$, são matrizes de coeficientes do tipo $k \times k$;
- $\boldsymbol{\varphi}$ é um vector de constantes de dimensão $k \times 1$;
- $\boldsymbol{\varepsilon}_t$ é um processo de ruído branco vectorial, $\boldsymbol{\varepsilon}_t \sim N(0, \Omega)$, com dimensão k ;
- Ω é a matriz de variâncias e covariâncias (definida positiva) ³⁸.

Para melhor compreender a equação (2.10), podemos representar o modelo VAR em que $k = 2$ e $p = 1$, VAR(1) (usualmente designado como VAR *simples*) do seguinte modo:

$$\mathbf{y}_t = \boldsymbol{\varphi} + \mathbf{A}_1 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2.11)$$

O que equivale a escrever:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Deste modo, os modelos VAR são modelos de regressão definidos por sistemas de equações que resultam numa combinação de séries temporais de sistemas de equações simultâneas. Assim, um modelo desta natureza trata todas as relações lineares existentes entre as variáveis *endógenas* e os valores passados das mesmas, permitindo a inclusão de variáveis *exógenas*.

Importa referir que o comportamento dos y 's depende das características das matrizes $\mathbf{A}_i, i = 1, \dots, p$. Com efeito, no que respeita à estabilidade dos modelos VAR, esta depende das soluções, λ , da equação

$$|\lambda^p \mathbf{I} - \lambda^{p-1} \mathbf{A}_1 - \dots - \lambda \mathbf{A}_{p-1} - \mathbf{A}_p| = 0$$

Especificamente, se todas as raízes estão no interior do círculo unitário, isto é, se $|\lambda| < 1$, o modelo é estável (estacionário), caso contrário, isto é, existe uma solução, λ , tal que $|\lambda| = 1$, o modelo é instável.

³⁸ Os ε 's não estão correlacionados em série, mas podem estar contemporaneamente correlacionados.

No que respeita ao processo de estimação do VAR, depois de analisadas quais as séries a considerar no modelo, o processo pode ser sintetizado nas seguintes etapas³⁹:

- estudar a estacionaridade, isto é, verificar a existência de raiz unitária, das séries em estudo ou por vezes é útil considerar os dados logaritmizados (ou os retornos dos dados);
- introduzir, caso se justifique, componentes sazonais (mais frequente em dados macroeconómicos);
- determinar o número de *lags* óptimo a considerar, o qual é escolhido a partir de critérios de informação/selecção;
- estimar os parâmetros do modelo utilizando o OLS;
- verificar a adequabilidade do modelo, analisando os pressupostos dos erros;

Percorridas estas etapas, caso se verifique que o modelo não é adequado, retomamos o processo caso contrário, estamos em condições de efectuar previsões, analisar a causalidade das séries, entre outros aspectos.

No entanto, apesar das vantagens reconhecidas nos modelos VAR, estes evidenciam várias limitações. Além de outros inconvenientes⁴⁰, os modelos VAR requerem que todas as componentes do modelo sejam estacionárias ou não cointegradas, com efeito, caso tal não se verifique são necessários adoptar procedimentos alternativos com vista à correcção do erro – MCE, surgindo assim os modelos VECM, nos quais é dada especial ênfase às relações de longo prazo. Com efeito, os modelos VECM distinguem-se dos modelos VAR pelo facto de apresentarem, como variáveis determinísticas, os MCE (sob a forma de vectores) correspondentes às relações de cointegração entre variáveis.

Procurando evitar uma apresentação densa, apresenta-se de seguida um encadeamento das ideias principais à compreensão dos aspectos em causa, procurando descrever já o que acontece no caso geral.

Partindo do VAR de primeira ordem, uma extensão para sistemas de ordem superior pode ser ilustrada com um sistema de segunda ordem definido por:

$$\mathbf{y}_t = \boldsymbol{\varphi} + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t \quad (2.12)$$

³⁹ Dada a forma sucinta com o processo é aqui descrito, para mais detalhes, entre outras referências possíveis, consultar Johnston e Dinardo (2001). No entanto, mais alguns detalhes sobre os procedimentos implementados serão ainda objecto de análise na parte prática do nosso trabalho.

⁴⁰ Mendes (2010) analisa outros inconvenientes dos modelos VAR, como por exemplo: modelos ateóricos, tal como os modelos ARMA; escolha adequada do número de *lags*; multiplicidade de parâmetros a estimar.

Se subtrairmos y_{t-1} a cada um dos membros, a equação (2.12) pode ser reescrita na forma:

$$\underbrace{y_t - y_{t-1}}_{\Delta y_t} = \varphi + (A_1 - I) y_{t-1} + A_2 y_{t-2} + \varepsilon_t \quad (2.13)$$

Pelo que, adicionando e subtraindo $(A_1 - I) y_{t-2}$ ao segundo membro da equação (2.13), obtemos

$$\Delta y_t = \varphi + (A_1 - I) \Delta y_{t-1} - (I - A_1 - A_2) y_{t-2} + \varepsilon_t \quad (2.14)$$

Ou seja, denotando $\Pi = I - A_1 - A_2$, uma reparametrização alternativa pode ser

$$\Delta y_t = \varphi - A_2 \Delta y_{t-1} - \Pi y_{t-1} + \varepsilon_t \quad (2.15)$$

Observe-se que, na reformulação (da primeira diferença) de um VAR(2), o segundo membro da equação apresenta um termo da primeira diferença desfasada.

Segundo este raciocínio, os modelos VECM surgem, assim, como uma reparametrização dos modelos VAR, sendo um modelo VECM(p) definido por:

$$\Delta y_t = \varphi + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-(p-1)} - \Pi y_{t-1} + \varepsilon_t \quad (2.16)$$

o que equivale a escrever:

$$\Delta y_t = \varphi + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} - \Pi y_{t-1} + \varepsilon_t \quad (2.17)$$

onde, para cada $i = 1, \dots, p$, as matrizes Γ_i são escritas em função das matrizes A e $\Pi = I - \sum_{i=1}^p A_i$. De acordo com a equação (2.17), o termo $\Pi = I - \sum_{i=1}^p A_i$ representa a informação de longo prazo e o termo $\Gamma_i = \sum_{j=i+1}^p A_j$ traduz a informação de curto prazo.

A matriz de informação de longo prazo, Π , pode ser escrita na forma $\Pi = \alpha\beta'$, onde α representa a velocidade de ajustamento ao desequilíbrio e β representa a matriz de coeficientes de longo prazo, ou seja os vectores cointegrantes r .

Assim, no que respeita à análise multivariada (modelos VAR e VECM), existem três situações específicas:

- utilização do modelo VAR em níveis, caso todas as variáveis y sejam estacionárias;

- utilização do modelo VAR em primeiras diferenças, caso todas as variáveis y sejam integradas de primeira ordem (não estacionárias) e não existam relações de cointegração;
- utilização do modelo VECM: caso todas as variáveis y sejam não estacionárias, mas integradas da mesma ordem e com pelo menos uma relação de cointegração.

Uma primeira consiste na estimação directa do sistema representado em (2.10) ou da reparametrização alternativa apresentada em (2.16), a qual só é apropriada se todos os valores próprios da matriz Π forem inferiores a um. Uma segunda abordagem é usada nos casos em que as variáveis y não são estacionárias, não impedindo tal facto que existam relações cointegrantes. Neste caso procede-se, primeiramente, à determinação da característica cointegrante⁴¹, r , para, posteriormente, estimar a equação (2.16) restringindo a matriz Π às r variáveis cointegrantes. Existem várias metodologias para tratar esta questão, sendo que a abordagem que têm merecido maior interesse (e a mais utilizada) é a de máxima verosimilhança apresentada em parte dos trabalhos desenvolvidos por Johansen, já referida na secção anterior.

Denotando por r e k a característica e a ordem da matriz Π , respectivamente, temos três possibilidades a considerar (Teorema da representação de Granger):

▪ $r = k$

Se cada raiz, λ , verificar $|\lambda| < 1$, a matriz Π tem k linhas (ou colunas) linearmente independentes, pelo que todas as variáveis y na equação (2.10) são estacionárias⁴². Neste caso, as estimativas OLS sem restrições da equação (2.10) ou da equação (2.16) produzem inferências idênticas dos parâmetros.

▪ $r < k$

Se existir uma raiz unitária com multiplicidade $k - r$ e se as restantes raízes forem inferiores a um, então a matriz Π tem um número de linhas (ou colunas) linearmente independentes inferior a k , sendo que as variáveis y não são estacionárias, mas são-o nas primeiras diferenças, ou nas diferenças de ordem superior⁴³. Assim, a matriz

⁴¹ Determinar a característica cointegrante, sendo $\Pi = \alpha\beta'$, equivale a determinar quantos vectores cointegrantes existem em β , ou seja, quantas colunas nulas existem em α (número de linhas/colunas linearmente independentes em Π).

⁴² Integradas de ordem zero, $I(0)$.

⁴³ Integradas de ordem p , $I(p)$, onde p denota a ordem onde se verifica a estacionariedade.

pode ser expressa em termos do produto de duas matrizes $(\alpha\beta')$ do tipo $k \times r$, ambas com característica r . Neste caso, que é o mais frequente, as variáveis estão cointegradas e o segundo membro da equação (2.16) contém r vectores cointegrantes.

▪ $r = 0$

No caso particular em que $I = \sum_{i=1}^p A_i$, temos $\Pi = 0$ (a matriz tem característica nula), pelo que não existe relação cointegrante no sistema. Neste caso, a equação (2.16) mostra que o VAR deveria ser especificado apenas em termos das primeiras diferenças das variáveis, isto é, pode usar-se um VAR em primeiras diferenças sobre os elementos de longo prazo.

No seguimento do exposto, um aspecto que merece particular análise (um dos temas centrais deste trabalho), está relacionado com a determinação da ordem de cointegração.

No que respeita à abordagem clássica para a determinação do(s) vector(es) cointegrante(s), consideramos neste trabalho o teste sugerido por Johansen (1988) que, além de permitir estimar todos os vectores cointegrantes, também permite testar restrições sobre os coeficientes dos respectivos vectores (Teste LR de proporcionalidade aplicado aos β 's e Teste de exogeneidade fraca aplicado aos α 's para averiguar a existência de causalidade ao longo prazo).

Especificamente, para testar se as raízes características estimadas ou os valores próprios da matriz Π são significativamente diferentes de zero, isto é, testar a hipótese da característica cointegrante ser quanto muito $r (< k)$, Johansen identifica dois testes estatísticos, sendo a estatística do traço a mais 'poderosa':

- a estatística do traço: $\lambda_{tr} = -T \sum_{i=r+1}^n \ln(1 - \lambda_i)$
- a estatística do valor próprio max: $\lambda_{max} = -T \ln(1 - \lambda_{r+1})$

Em termos de hipóteses nula e alternativa, cada um dos testes caracteriza-se do seguinte modo:

ESTATÍSTICA DO TRAÇO	ESTATÍSTICA DO VALOR PRÓPRIO MAX
H_0 : Existem pelo menos r vectores cointegrantes	H_0 : Existem r vectores cointegrantes
vs	vs
H_1 : Existem $k > r$ vectores cointegrantes	H_1 : Existem $r + 1$ vectores cointegrantes

O facto das distribuições das estatísticas do teste não serem *standard* obriga a que os valores críticos sejam obtidos por simulação. Osterwald-Lenum (*apud* Johston e Dinardo,

2001), desenvolveu um dos mais completos trabalhos apresentando um conjunto de valores críticos para VAR até onze variáveis, os quais são agrupados em tabelas distintas, diferindo estas em função das múltiplas possibilidades de especificação do VAR (inclusão de termos constantes e tendências temporais quer nas equações VAR como nas equações cointegrantes). Com efeito, para levar a cabo o teste da característica de cointegração é necessário escolher a especificação que parece mais plausível para os dados em análise.

Procurando sintetizar algumas ideias, apresentam-se no esquema seguinte algumas das etapas a percorrer na implementação a metodologia de Johansen.

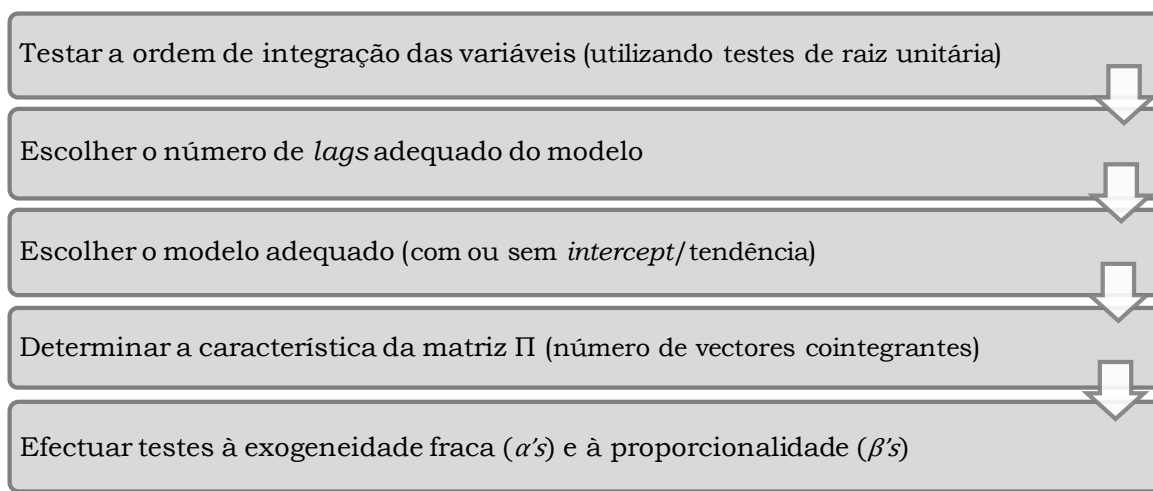


FIGURA 8 – Esquema de implementação da Metodologia de Johansen

Para finalizar esta secção, é importante sublinhar que apesar das potencialidades reconhecidas, alguns casos poderão não ser resolúveis usando a metodologia de Johansen. São exemplo as fragilidades verificadas perante alguns cenários hipotéticos: (i) o caso em que as duas estatísticas (*traço* e *valor próprio Max*) não são consensuais; (ii) o facto de se estar perante sistemas de grande dimensão ou ser necessário considerar uma ordem elevada de *lags* para captar as dinâmicas de curto prazo; (iii) situações de amostras com reduzida dimensão, pois as frequências empíricas de rejeição são em excesso em relação ao nível nominal; (iv) dinâmicos e estruturas não-lineares subjacentes a séries em estudo. Neste seguimento, várias são as referências na literatura sobre o tema, ao facto de que o teste de razão de verosimilhança (desenvolvido por Johansen), para a determinação da ordem de integração, pode ser pouco credível.

Precisamente, face a estas e outras limitações da abordagem clássica aos modelos VAR (e conseqüente extensão dos mesmos – modelos VECM), na secção seguinte serão apresentados os modelos VAR segundo uma abordagem bayesiana, os quais vêm dar

resposta a alguns dos problemas anteriormente intratáveis e introduzir abordagens ou modelos mais flexíveis, estando um dos aspectos precisamente relacionado com uma abordagem alternativa à determinação da ordem de cointegração. Isto porque, em certo sentido, a análise bayesiana da cointegração é directa e imediata. A equação do modelo VAR ou VECM, juntamente com os seus pressupostos, definem a função verosimilhança, sendo possível combinar esta função com a distribuição *a priori* e fazer a inferência bayesiana com a distribuição *a posteriori* resultante.

2.4. MODELOS ECONOMETRICOS E COINTEGRAÇÃO: ABORDAGEM BAYESIANA

Determinar a ordem de cointegração de um sistema torna-se, cada vez mais, um aspecto fundamental da investigação aplicada em macroeconomia e finanças.

Face a considerações já tecidas, as vantagens da aplicação dos pressupostos bayesianos, quando se pretende estudar modelos VAR, resultam, em parte, destes permitirem combinar a informação amostral com a informação admitida *a priori* através do Teorema de Bayes. Por exemplo, para uma estimação mais eficiente do modelo VAR é requerida uma grande quantidade de informação amostral, não sendo possível aumentar a dimensão desta, a solução passa por incorporar informação *a priori*. Outras vantagens são, ainda, reconhecidas por permitirem contornar alguns problemas intratáveis decorrentes da aplicação de metodologias clássicas.

De um modo sucinto, no que respeita aos modelos VAR, a aplicabilidade da abordagem bayesiana consiste em impor restrições *a priori* sobre as distribuições dos parâmetros do modelo VAR, sendo que os parâmetros do modelo são obtidos combinando as distribuições *a priori* com a informação amostral.

Porém, dada a extensão dos assuntos em causa e tendo em conta o tema central do nosso trabalho, nesta secção apenas será dada ênfase a aplicabilidade da abordagem bayesiana na cointegração: modelos VAR cointegrados (ou modelos BVECM) e respectiva ordem de cointegração.

2.4.1. CONSIDERAÇÕES GERAIS NA ABORDAGEM BAYESIANA

No seguimento das limitações resultantes da abordagem usual na análise de cointegração, seguindo a abordagem de máxima verosimilhança sugerida por Johansen, a que nos referimos na secção anterior, vários trabalhos foram desenvolvidos no sentido de

dar resposta às limitações identificadas, apresentando para o efeito uma abordagem bayesiana para a cointegração. São exemplo desses desenvolvimentos os trabalhos desenvolvidos por Kleibergen e van Dijk (1994), Bauwens e Lubrano (1996), Geweke (1996), Bauwens e Giot (1998), Kleibergen e Paap (2002), Strachan (2003), Strachan e Inder (2004) e Villani (2005), citados em Warne (2006), sendo que todos estes estudos têm em comum o pressuposto de que o modelo cointegrado VAR tem resíduos normalmente distribuídos e, conseqüentemente, têm a mesma função de verosimilhança do método clássico de Johansen.

De entre as principais vantagens práticas subjacentes à abordagem bayesiana, Warne (2006) destaca dois factos: por um lado, as distribuições de probabilidades, para cada conjunto de parâmetros, são sempre válidas, independentemente do tamanho da amostra; por outro lado, é possível fazer inferências sobre a ordem de cointegração e outras restrições aos parâmetros do modelo.

Uma questão crucial que tem bastante interesse e importância empírica, está relacionada com o problema de identificação, ou seja, caso em que a matriz Π tem característica reduzida. A identificação global pode ser observada, facilmente, pelo facto da matriz $\Pi = \alpha\beta'$ e a matriz $\Pi = \alpha A A^{-1} \beta'$ serem idênticas para qualquer matriz A regular (invertível). Esta indeterminação pode ser resolvida considerando a normalização linear. Porém, a normalização traz, também, desvantagens ao modelo, pois, por exemplo, para uma distribuição *a priori* ‘não-informativa’ comum, podemos obter uma distribuição *a posteriori* imprópria, o que vai impedir validar a inferência estatística. Todos os trabalhos científicos mais recentes sobre cointegração debruçam-se sobre a resolução ou melhoria destes problemas.

Os trabalhos pioneiros na cointegração bayesiana estão relacionados com os modelos BVAR que ignoram a determinação da ordem de cointegração pela não utilização dos modelos BVECM. São exemplo os trabalhos desenvolvidos por DeJong (1992), Koop (1991) e Koop (1994). Neste caso, a computação bayesiana fica bastante mais simples, pela não inclusão da característica reduzida da matriz de cointegração na distribuição *a priori*.

DeJong (1992) utiliza um método VAR com uma distribuição *a priori* não-informativa para os parâmetros do modelo. Depois, implementa-se uma integração de Monte Carlo para obter a distribuição *a posteriori*, que, finalmente, vai ser utilizada para obter as raízes da representação VAR. A ordem de cointegração está directamente relacionada com o número de raízes não estacionárias (sobre o círculo unitário) da representação VAR, cuja probabilidade é calculada utilizando os *outputs* da integração de Monte Carlo.

Mais tarde, Kleibergen e van Dijk (1994), Bauwens e Lubrano (1996) e Geweke (1996), consideram uma nova abordagem da cointegração, condicionando um número dado de vectores cointegrantes e trabalhando, directamente, com as matrizes α e β . Por outras palavras, a natureza de característica reduzida da matriz Π , é directamente imposta no algoritmo de simulação *a posteriori*. Isto permite a estimação e a inferência bayesiana sobre α e β (e outros parâmetros do modelo) para dado r . Considerando inferência *a posteriori* para cada valor possível de r e utilizando métodos standard de comparação de modelos, podemos, finalmente, seleccionar a ordem de cointegração adequada.

Observa-se, também, que a não-linearidade obtida pela implementação do modelo VECM vem eliminada pelo condicionamento sobre o vector de cointegração, β .

Sendo assim, pela escolha certa de uma distribuição *a priori* informativa, podemos aplicar uma análise standard de tipo VAR. Mais, para especificações particulares, a distribuição *a posteriori* de β condicionada a α , Σ , Ω tem uma distribuição standard, o que implica que pode ser utilizado um método de simulação *a posteriori* de tipo Gibbs.

Não obstante destas considerações, centremos-nos, daqui em diante, nos trabalhos desenvolvidos por Villani (2005) e Warne (2006).

Tal como sublinhado por Warne (2006), um passo crucial na análise bayesiana consiste, precisamente, na escolha da distribuição *a priori*. Na literatura sobre cointegração/modelos BVAR cointegrados, várias têm sido as abordagens e as escolhas sugeridas para esta distribuição (para uma revisão dessas escolhas consulte Koop *et al.* (2006)), cuja opção mais frequente é a que considera *prioris* subjectivas contendo, apenas, uma pequena informação, isto é, *prioris* amplamente dominadas pelos dados.

No estudo desenvolvido por Villani (2005) o principal enfoque não é averiguar se a distribuição *a priori* é ou não informativa, mas sim propor uma análise bayesiana onde a distribuição *a priori* utilizada seja apelativa e de fácil implementação. Segundo o mesmo autor, a distribuição *priori* deve considerar alguns aspectos: (i) o número de parâmetros nos modelos de cointegração é normalmente muito elevado e como tal a estimação de tais parâmetros da *priori* não é realista em espaços de elevada dimensão. Assim, um dos requisitos preponderantes é uma *priori* com um número reduzido de parâmetros e cuja interpretação destes seja clara; (ii) a *priori* deve ser transparente no sentido de ser facilmente perceptível o tipo de informação que a mesma transmite; (iii) a obtenção da *posteriori* (a partir da *priori*) deve ser simples, permitindo ser realizada de forma ‘rotineira’; (iv) a computação da distribuição *a posteriori* da ordem de cointegração deve ser viável.

Neste seguimento, Warne (2006) procurou estender a abordagem bayesiana a duas dimensões diferentes. Por um lado, permitir, para uma distribuição *a priori* informativa, parâmetros que sejam variáveis endógenas desfasadas, ou seja, parâmetros que representam a dinâmica de curto prazo. Por outro, basear na função de verosimilhança marginal identidade⁴⁴ a aproximação para o cálculo das probabilidades da ordem da *posterior*.

Apresentada uma visão geral, as considerações tecidas carecem de uma formalização, para uma compreensão mais consistente dos procedimentos práticos que nos propomos desenvolver. Com efeito, toda a apresentação e análise feitas nas páginas subsequentes, tem por base os trabalhos de Warne (2006) e Villani (2005).

Procurando estabelecer uma ponte com a abordagem ao modelo VAR cointegrado feita na Secção 2.3 (modelos VECM), retomemos sucintamente alguns conceitos e modelos analisados.

Considere-se o processo p -dimensional, x_t , representativo de um modelo VAR cointegrado com r relações de longo prazo:

$$\Delta x_t = \Phi D_t + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Pi x_{t-1} + \varepsilon_t, \quad t = 1, \dots, T \quad (2.18)$$

onde, à semelhança do visto anteriormente:

- x_t é um vector coluna com p variáveis;
- $\Gamma_i, i = 1, \dots, k - 1$, são matrizes de coeficientes do tipo $p \times p$;
- $\Pi = \alpha\beta'$, onde α e β são matrizes de coeficientes do tipo $p \times r$, para um certo $r = 0, 1, \dots, p$;
- ΦD_t é um vector de constantes de dimensão $k \times 1$, onde D_t ($d \times 1$) é um vector de constantes, tendência, *dummies* sazonais, ou outras variáveis (deterministas ou exógena);⁴⁵
- ε_t é um processo de ruído branco vectorial, $\varepsilon_t \sim N(0, \Omega)$;
- Ω é a matriz de variâncias e covariâncias (definida positiva).

⁴⁴ Esta identidade afirma que a verosimilhança marginal é dada pelo produto da densidade dos dados condicionados aos parâmetros com a densidade *a priori*, dividido pela densidade *a posteriori*. Todas estas densidades têm de ser avaliadas no domínio destes parâmetros.

⁴⁵ Podem existir modelos onde o vector D_t não tenha expressão.

A equação (2.18) pode ser reescrita na forma:

$$Z_{0t} = \Phi D_t + \Gamma Z_{2t} + \alpha\beta'Z_{1t} + \varepsilon_t \quad (2.19)$$

onde $\Gamma = [\Gamma_1, \dots, \Gamma_{k-1}]$, $Z_{0t} = \Delta x_t$, $Z_{1t} = x_{t-1}$ e $Z_{2t} = [\Delta x'_{t-1} \dots \Delta x'_{t-k+1}]'$. Dada a utilidade em escrever o modelo de uma forma mais sucinta, a equação (2.19) pode ser reescrita como:

$$Z_0 = \Phi D + \Gamma Z_2 + \alpha\beta'Z_1 + \varepsilon \quad (2.20)$$

Neste caso, $Z_0 = [Z_{01}, \dots, Z_{0T}]$ é uma matriz $(p \times T)$ e as matrizes D, Z_1, Z_2 e ε são definidas de forma idêntica tendo T colunas. Por forma a abreviar a notação referente a todos os dados disponíveis, considere-se, daqui em diante, $\mathcal{D} = \{D, Z_0, Z_1, Z_2\}$.

2.4.2. DISTRIBUIÇÃO A PRIORI

Um passo crucial na análise bayesiana consiste na escolha dos parâmetros da distribuição *a priori*. Contudo, essa escolha não é de todo linear pois, duas características importantes do VAR cointegrado fazem com que *as prioris* estandardizadas sejam inadequadas na análise da cointegração. Primeiro, as restrições de característica reduzida impostas sobre a matriz $\Pi = \alpha\beta'$, podem conduzir à não linearidade de modelos que eram lineares. Segundo, estes resultados de não linearidade dos parâmetros (α, β) são não-identificados, na medida em que apenas o espaço gerado pelas colunas de β pode ser unicamente determinado.

Consequentemente os vectores de cointegração têm de ser restritos de forma que todos os parâmetros sejam identificados.

Consideremos uma matriz C , com dimensão $p \times r$ onde r representa a ordem de cointegração dos sistema e a característica dessa matriz (*rank*), e onde $\bar{C} = C(C'C)^{-1}$. As restrições de identificação linear sobre β (normalização linear) podem ser escritas na forma $\beta_c = \beta(\bar{C}'\beta)^{-1}$ ou, equivalentemente:

$$\beta_c = C + C_\perp \Psi \quad (2.21)$$

onde, $\Psi = \bar{C}'_\perp \beta(\bar{C}'\beta)^{-1}$.

Por exemplo, sendo C dado pelas primeiras r colunas de I_p e β identificada pela equação (2.21), temos $\alpha_c = \alpha\beta'\bar{C}$ é tal que $\alpha_c\beta'_c = \alpha\beta' = \Pi$.

Warne (2006) sublinha que as restrições lineares identificadas em (2.21) envolvem o pressuposto que $C'\beta$ tem característica completa r . Se este pressuposto não se verificar, então as restrições não são válidas⁴⁶.

Quando α tem ordem reduzida (ou igual a zero), a distribuição *a posteriori* para os parâmetros livres de β condicionados a α é igual à sua *a priori*, uma vez que os parâmetros livres de β não são considerados na densidade dos dados. Logo, impor restrições de exogeneidade fraca em α , pode causar problemas na estimação da distribuição *a posteriori*.

Para resolver alguns destes problemas foram sugeridas três aproximações, sendo a mais promissora e natural aquela que considera *a priori* sobre o espaço de cointegração (em vez da ordem de cointegração), o qual é baseado num conjunto de todas as matrizes de dimensão $p \times r$. Para mais detalhes sobre estes aspectos consultar Warne (2006) e Villani (2005). O mesmo autor remete ainda para os artigos de Strachan e van Dijk (2003) e Strachan e Inder (2004), os quais analisam uma estratégia alternativa para evitar as restrições que identificam a linearidade de β .

Antecedendo à análise da distribuição *a priori*, vejamos as seguintes definições relativas à distribuição *t* e da distribuição *Wishart* invertida.

DEFINIÇÃO 2.2

Seja B , matriz aleatória de dimensões $m \times s$. As entradas da matriz B têm distribuição *t*, ou, simplesmente, B é uma matriz *t* com parâmetros μ, P, Q e n , com $n \geq 0$, $\mu \in \mathbb{R}^{m \times s}$, P e Q matrizes de ordem m e s , respectivamente, em que Q é definida positiva, isto é, $B \sim t_{m \times s}(\mu, P, Q, n)$ se a função densidade é dada por:

$$p(B) = \frac{\Gamma_s(n+m+s)|P|^{s/2}}{\Gamma_s(n+s)\pi^{ms/2}|P|^{s/2}} |I_s + Q^{-1}(B - \mu)'P(B - \mu)|^{-(n+m+s)/2} \quad 47$$

DEFINIÇÃO 2.3

Uma matriz aleatória, Σ , definida positiva de ordem m tem uma distribuição *Wishart* invertida com parâmetros S (matriz definida positiva) e $n \geq m$, isto é, $\Sigma \sim IW_m(S, n)$ se a função densidade é dada por:

$$p(\Sigma) = \frac{|S|^{n/2}}{\Gamma_m(n)\pi^{m(m-1)/4} 2^{nm/2}} |\Sigma|^{-(n+m+1)/2} \exp\left(-\frac{1}{2} \text{tr} [\Sigma^{-1}S]\right)$$

⁴⁶ Porém, independentemente da validade, ou não, das restrições, o problema da selecção *a priori* para os vectores cointegrados não é resolvido (ver Strachan e van Dijk, 2004, citado em Warne, 2006)

⁴⁷ $\Gamma(\cdot)$ denota a função Gama, tendo-se $\Gamma_b(a) = \prod_{i=1}^b \Gamma\left(\frac{a-i+1}{2}\right)$, $a, b \in \mathbb{N}$ com $a \geq b$. Se $b = 0$, então $\Gamma_0(a) = 1$ para $a \geq 0$.

As propriedades relativas às distribuições acima referidas, podem ser consultadas em Bauwens, *et al.* (1999), Box and Tiao (1973), citados em Warne (2006).

A distribuição *a priori* usada para a análise da ordem de cointegração é baseada na distribuição *a priori* de referência desenvolvida por Villani (2005). A distribuição conjunta é decomposta como:

$$p(\alpha, \beta, \Phi, \Gamma, \Omega, r) = p(\alpha, \beta, \Phi, \Gamma, \Omega | r) p(r) \quad (2.22)$$

onde deve ser utilizada uma *priori* uniforme para a ordem de cointegração, ou seja, $p(r) = 1/(p + 1)$, para todo $r \in \{0, 1, \dots, p\}$.

A densidade *a priori* de $p(\alpha, \beta, \Phi, \Gamma, \Omega)$, condicionada à ordem de cointegração, é dada por:

$$p(\alpha, \beta, \Phi, \Gamma, \Omega | r) = c_r |\Omega|^{-(p+q+r+1)/2} \exp \left(-\frac{1}{2} \text{tr} [\Omega^{-1} (A + (1/\lambda_\alpha^2) \alpha \beta' \beta \alpha')] \right) p(\Gamma | \Omega) \quad (2.23)$$

onde os híper-parâmetros $\lambda_\alpha > 0$, $q \geq p$ e A (matriz definida positiva de ordem p) são especificados pelo utilizador. Ainda, na equação (2.23), c_r denota a constante normalizadora dada por:

$$c_r = v_r |A|^{q/2} \frac{\Gamma_r(p)}{\Gamma_p(q) \Gamma_r(r)} \frac{2^{-qp/2} \pi^{-p(p-1)/4}}{(2\pi \lambda_\alpha^2)^{pr/2} \pi^{(p-r)r/2}}$$

onde v_r depende da normalização escolhida para β quando $r \in \{1, \dots, p-1\}$, enquanto que $v_0 = v_p = 1$. Especificamente, para $\beta = \beta_c I$, temos $v_r = |c'_\perp c_\perp|^{r/2} |c' c|^{r/2}$.

Evitando tornar densa esta apresentação, para mais detalhes sobre a equação (2.23) consultar Warne (2006).

Observe-se ainda que, partindo do estudo desenvolvido por Villani (2005), o qual considera o caso particular $p(\Gamma | \Omega) = 1$ na análise e apresentação de alguns resultados, Warne (2006) refere alguns desenvolvimentos teóricos, tamando para o efeito:

$$p(\Gamma | \Omega) = (2\pi)^{-\frac{p^2(k-1)}{2}} |\Sigma_\Gamma|^{-\frac{p}{2}} |\Omega|^{-\frac{p(k-1)}{2}} \exp \left(-\frac{1}{2} \text{tr} [\Omega^{-1} \Gamma \Sigma_r^{-1} \Gamma'] \right) \quad (2.24)$$

Neste seguimento, os resultados analisados na secção seguinte terão sempre em linha de conta estes dois casos.

2.4.3. DISTRIBUIÇÃO A POSTERIORI

Sendo a distribuição *a posteriori* conjunta dos parâmetros do modelo VAR cointegrado intratável, impõe-se a busca de soluções. Apresenta-se de seguida resultados necessários para a avaliação numérica da distribuição *a posteriori* via amostragem *Gibbs*. Esta amostragem é simulada a partir da distribuição *a posteriori* conjunta dos parâmetros do modelo gerando, de forma iterativa, resultados a partir da distribuição *a posteriori* completa.⁴⁸

No seguimento do exposto na Secção 1.4.3.2 do nosso trabalho, retomemos algumas considerações procurando estabelecer uma ponte com os aspectos a apresentar.

Dado o conjunto completo de parâmetros do modelo, θ , onde θ_j é o j -ésimo grupo de parâmetros, com $\theta = (\theta_1, \dots, \theta_g)$, e onde g é o número de grupos, aq distribuição *a posteriori* condicionada completa⁴⁹ para θ_j é denotada por

$$p(\theta_j | \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_g, \mathcal{D}, M_m)$$

onde M_m representa o modelo m . Tal como visto anteriormente, a amostragem de *Gibbs* começa com $\theta^{(0)}$, para alguns valores do suporte da distribuição *a posteriori* conjunta, enquanto que $\theta^{(i)}$ é a i -ésima simulação da amostragem de *Gibbs*. A i -ésima simulação de θ é obtida pelo conjunto de simulações das *posteriors* completas condicionadas a θ_j , $j = 1, \dots, g$, onde

$$\theta_j^{(i)} \sim p(\theta_j | \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_g^{(i-1)}, \mathcal{D}, M_m), \quad j = 1, \dots, g; \quad i = 1, \dots, G$$

tendo-se $\theta^{(i)} = (\theta_1^{(i)}, \dots, \theta_g^{(i)})$.

Assim, sendo que $\theta^{(i-1)}$ é uma simulação da distribuição *a posteriori*, então, por definição de densidade condicionada, para cada passo j , temos:

$$(\theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_j^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_g^{(i-1)}) \sim p(\theta | \mathcal{D}, M_m)$$

Pelo que, $\theta^{(i)}$ é também uma simulação da distribuição conjunta *a posteriori*.

⁴⁸ Ver Casella e George (1992), Tierney (1994), e Geweke (1999), citado em Warne (2006).

⁴⁹ A distribuição *a posteriori* condicionada completa é uma distribuição *a posteriori* para um grupo de parâmetros condicionados aos dados e aos parâmetros remanescentes

Deste modo, os resultados da amostragem de *Gibbs* é uma sequência das simulações $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(G)}$ da distribuição conjunta *a posteriori* de θ . Devido às propriedades das cadeias de Markov da amostragem de *Gibbs*, as simulações não são independentes mas sob certas condições⁵⁰, as quais são satisfeitas aqui, pode provar-se que:

$$f(\theta^{(i)}) \xrightarrow{d} p[f(\theta) | \mathcal{D}, M_m], \text{ e } \frac{1}{G} \sum_{i=1}^G f(\theta^{(i)}) \xrightarrow{q.c.} \mathbb{E}[f(\theta) | \mathcal{D}, M_m]$$

Por isso, dadas as simulações da *posteriori* de θ , a distribuição *a posteriori* de qualquer função dos parâmetros do modelo, tal como a função de impulso resposta, estão imediatamente disponíveis aplicando a função f apropriada a cada simulação *a posteriori* e usando algum estimador da densidade.

Com $\theta = (\alpha, \psi, \Phi, \Gamma, \Omega)$, a proposição seguinte dá-nos *a posteriori* condicionada completa dos cinco grupos de parâmetros em θ dado r .

PROPOSIÇÃO 2.1 (Warne, 2006)

As distribuições a posteriori condicionadas completas são:

$$\Omega | \alpha, \beta, \Phi, \Gamma, \mathcal{D}, r \sim IW_p \left(\varepsilon \varepsilon' + A + \frac{1}{\lambda_\alpha^2} \alpha \beta' \beta \alpha' + \Gamma \Sigma_r^{-1} \Gamma', T + q + r + m \right)$$

onde $\varepsilon = Z_0 - \Phi D - \Gamma Z_2 - \alpha \beta' Z_1$.

Se $p(\Gamma | \Omega) = 1$, então $m = 0$ e $\Sigma_r^{-1} = 0$. Caso $p(\Gamma | \Omega)$ seja dado pela equação (2.24), então $m = p(k - 1)$.

Tem-se:

(i) $\text{vec}(\Phi) | \alpha, \beta, \Gamma, \Omega, \mathcal{D}, r \sim N_{pd}(\bar{\mu}_\Phi, \bar{\Sigma}_\Phi)$, onde:

$$\bar{\mu}_\Phi = \text{vec} \left([Z_0 - \Gamma Z_2 - \alpha \beta' Z_1] D' (DD')^{-1} \right) \text{ e } \bar{\Sigma}_\Phi = [(DD')^{-1} \otimes \Omega].$$

(ii) $\text{vec}(\Gamma) | \alpha, \beta, \Phi, \Omega, \mathcal{D}, r \sim N_{p^2(k-1)}(\bar{\mu}_\Gamma, \bar{\Sigma}_\Gamma)$, onde:

$$\bar{\mu}_\Gamma = \text{vec}([Z_0 - \Phi D - \alpha \beta' Z_1] Z_2' [Z_2 Z_2' + \Sigma_r^{-1}]^{-1}) \text{ e } \bar{\Sigma}_\Gamma = [(Z_2 Z_2' + \Sigma_r^{-1})^{-1} \otimes \Omega].$$

⁵⁰ Recordar secção 1.3., onde foram apresentados, sucintamente, alguns resultados relativos a cadeias de Markov, em particular relacionados com a convergência. Para mais detalhes, sobre estes factos e resultado teóricos associados, ver Tierney (1994).

(iii) $\text{vec}(\alpha)|\beta, \Phi, \Gamma, \Omega, \mathcal{D}, r \sim N_{pr}(\bar{\mu}_\alpha, \bar{\Sigma}_\alpha)$, onde:

$$\bar{\mu}_\alpha = \text{vec} \left([Z_0 - \Phi D - \Gamma Z_2] Z_1' \beta \left[\beta' \left(Z_1 Z_1' + \frac{1}{\lambda_\alpha^2} I_p \right) \beta \right]^{-1} \right) \text{ e}$$

$$\bar{\Sigma}_\alpha = \left[\left[\beta' \left(Z_1 Z_1' + \frac{1}{\lambda_\alpha^2} I_p \right) \beta \right]^{-1} \otimes \Omega \right].$$

(iv) $\psi|\alpha, \Phi, \Gamma, \Omega, \mathcal{D}, r \sim N_s(\bar{\mu}_\psi, \bar{\Sigma}_\psi)$, onde:

$$\bar{\mu}_\psi = \bar{\Sigma}_\psi H' \bar{\Sigma}_\beta^{-1} (\bar{\mu}_\beta - h) \text{ e } \bar{\Sigma}_\psi = (H' \bar{\Sigma}_\beta^{-1} H)^{-1}, \text{ em que } \bar{\mu}_\beta \text{ e } \bar{\Sigma}_\beta^{-1} \text{ são definidos por: } \bar{\mu}_\beta = \bar{\Sigma}_\beta \text{vec}(Z_1 [Z_0 - \Phi D - \Gamma Z_1]' \Omega^{-1} \alpha) \text{ e } \bar{\Sigma}_\beta^{-1} = \left[\alpha' \Omega^{-1} \alpha \otimes \left(Z_1 Z_1' + \frac{1}{\lambda_\alpha^2} I_p \right) \right].$$

O resultado seguinte dá-nos a *posteriori* condicionada marginal, sendo a *priori* informativa dada pela equação (2.24).

PROPOSIÇÃO 2.2 (Adaptado de Warne, 2006)

Suponhamos que $\beta = \beta_c$, e $\alpha = \alpha_c$, então

$$\alpha|\beta, \mathcal{D}, r \sim t_{p \times r}[\hat{\alpha}, (A + Z_0 N(Z_0 - \hat{\alpha} \beta' Z_1)')^{-1}, (\beta' C_1 \beta)^{-1}, T + q - (pn + d)]$$

onde, $\hat{\alpha} = Z_0 N Z_1' \beta (\beta' C_1 \beta)^{-1}$ e $C_1 = Z_1 N Z_1' + \frac{1}{\lambda_\alpha^2} I_p$.

Além disso,

$$\Psi|\alpha, \mathcal{D}, r \sim t_{(p-r) \times r}[\hat{\Psi}, (G_3 - G_2' G_1^{-1} G_2)^{-1}, C_3, T + q + r - (pn + d)]$$

onde:

- $\hat{\Psi} = \hat{\beta}_2 + G_2' - G_1^{-1} (I_r - \hat{\beta}_1)$
- $C_3 = (I_r - \hat{\beta}_1)' G_1^{-1} (I_r - \hat{\beta}_1) + (\alpha' S^{-1} \alpha)^{-1}$
- $\hat{\beta} = K \hat{\Pi}' S^{-1} \alpha (\alpha' S^{-1} \alpha)^{-1}$
- $\hat{\Pi} = Z_0 N Z_1' C_1^{-1}$
- $K = [\bar{c} \ \bar{c}_\perp]'$
- $S = A + Z_0 N Z_0' - Z_0 N Z_1' C_1^{-1} Z_1 N Z_0'$

$$\bullet \quad K C_1^{-1} K' + K \hat{\Pi}' S^{-1} \hat{\Pi} K' - \hat{\beta} \alpha' S^{-1} \alpha \hat{\beta}' = \begin{bmatrix} \underbrace{G_1}_{r \times r} & \underbrace{G_2}_{r \times (p-r)} \\ \underbrace{G_2'}_{(p-r) \times r} & \underbrace{G_3}_{(p-r) \times (p-r)} \end{bmatrix}$$

onde $\hat{\beta}_1$ e $\hat{\beta}_2$, são matrizes do tipo $(r \times r)$ e $((p-r) \times r)$, respectivamente, tais que $\hat{\beta} = [\hat{\beta}_1' \ \hat{\beta}_2']'$. A matriz de integração N , de ordem T (com respeito a Φ e Γ) e o inteiro n dependem da distribuição a priori de Γ .

Tem-se:

(i) Se $p(\Gamma|\Omega) = 1$, então,

$$N = M_z, \quad M_z = I_T - Z'(ZZ')^{-1}Z, \quad Z = [D' \ Z_2']', \quad n = k.$$

(ii) Se $p(\Gamma|\Omega)$ é dada pela equação (2.24), então,

$$N = M_D - M_D Z_2' (Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}) Z_2 M_D, \quad M_D = I_T - D'(DD')^{-1}D, \quad n = 1.$$

No que respeita à probabilidade da *a posteriori* da ordem de cointegração, esta pode ser determinada a partir da Regra de Bayes:

$$p(r|\mathcal{D}) = \frac{p(\mathcal{D}|r) p(r)}{\sum_{i=0}^p p(\mathcal{D}|i) p(i)} \quad (2.25)$$

Além disso, dado que as verosimilhanças marginais, $p(\mathcal{D}|r)$, estão disponíveis, a probabilidade da *a posteriori* da ordem de cointegração resulta directamente de (2.25).

Resta, agora, analisar as verosimilhanças marginais. Para os casos particulares em que $r = 0$ ou $r = p$, Warne (2006), no seguimento do trabalho de Villani (2005) (que define as respectivas verosimilhanças para o caso específico $p(\Gamma|\Omega) = 1$), apresenta uma expressão geral para as verosimilhanças marginais generalizando assim o resultado apresentado por Villani (2005, Teorema 5.1).

PROPOSIÇÃO 2.3 (Adaptado de Warne, 2006)

As verosimilhanças marginais para $r = 0$ e $r = p$ são dadas por:

$$p(\mathcal{D}|r = 0) = k_1 \Gamma_p(T + q - p(n - 1) - d) |A + Z_0 N Z_0'|^{-(T+q-p(n-1)-d)/2}$$

e

$$p(\mathcal{D}|r = p) = k_1 \Gamma_p(T + q - p(n - 1) - d) \lambda_\alpha^{-p^2} |C_1|^{-p/2} |S|^{-(T+q-p(n-1)-d)/2}$$

A constante multiplicativa é definida por:

$$k_1 = \frac{|A|^{q/2}}{\pi^{(T-p(n-1)-d)p/2} z^{p/2} \Gamma_p(q)}$$

onde:

(i) $z = |ZZ'|$ se $p(\Gamma|\Omega) = 1$;

(ii) $z = |\Sigma_\Gamma| |DD'| |Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}|$ se $p(\Gamma|\Omega)$ é dada pela equação (2.24).

Nos restantes casos em que a ordem de integração $r \in \{1, \dots, p-1\}$, a verosimilhança marginal pode ser calculada usando *a priori* definida em (2.24). Villani (2005) sugere três métodos para determinação de $p(\mathcal{D}|r)$: integração de Monte Carlo, distribuição de interesse e verosimilhança marginal identidade básica.

Neste contexto, a verosimilhança marginal identidade pode ser representada por:

$$p(\mathcal{D}|r) = \frac{p(\mathcal{D}|\alpha, \Psi, r) p(\alpha, \Psi|r)}{p(\Psi|\alpha, \mathcal{D}, r)p(\alpha|\mathcal{D}, r)} \quad (2.26)$$

Chib (1995), citado em Warne (2006), sugere usar esta identidade em combinação com a amostragem de *Gibbs* para calcular a verosimilhança marginal.

Para finalizar esta apresentação, analisemos a equação (2.26). No que respeita ao numerador da equação, este pode ser determinado analiticamente (Villani, 2005, Lema 5.3), tendo-se:

$$p(\mathcal{D}|\alpha, \Psi, r) p(\alpha, \Psi|r) = \\ = k_1 v_r \frac{\Gamma_p(T + q + r - p(n-1) - d) \Gamma_r(p)}{\Gamma_r(r) \pi^{(2p-r)r/2} \lambda_\alpha^{pr}} \times \left| A + \frac{1}{\lambda_\alpha^2} \alpha \beta' \beta \alpha' + W N W' \right|^{-(T+q+r-p(n-1)-d)/2}$$

onde $W = Z_0 - \alpha \beta' Z_1$ e os parâmetros N e n são os definidos na Proposição 2.2.

Quanto denominador, se $p(\Psi|\alpha, \mathcal{D}, r)$ é dado na Proposição 2.2, já o termo $p(\alpha|\mathcal{D}, r)$ não tem uma forma fechada específica. Contudo, no ponto $\alpha = \tilde{\alpha}$, o termo pode ser estimado, a partir da distribuição *a posteriori* $\Psi^{(1)}, \dots, \Psi^{(G)}$ de Ψ (que pode ser simulada da distribuição condicionada completa de ψ , identificada na Proposição 1.1, ou da distribuição condicionada marginal de Ψ , identificada na Proposição 2.2), sendo dado por:

$$\hat{p}(\tilde{\alpha}|\mathcal{D}, r) = \frac{1}{G} \sum_{i=1}^G p(\tilde{\alpha}|\Psi^{(i)}, \mathcal{D}, r)$$

onde $p(\Psi|\alpha, \mathcal{D}, r)$ é dado pela Proposição 2.2, cuja precisão de $\hat{p}(\tilde{\alpha}|\mathcal{D}, r)$ pode ser avaliada através do seu erro numérico padrão⁵¹.

⁵¹ Para mais detalhes, consultar Warne e Chib (1995), citado em Warne (2006).

3. ANÁLISE DESCRITIVA DAS SÉRIES TEMPORAIS

No presente capítulo, procurar-se-á fazer uma descrição e análise das séries temporais que servirão de base à implementação prática de aspectos teóricos apresentados. Para o efeito, fará sentido uma apresentação prévia das séries em estudo e consequente análise descritiva das mesmas, onde, para além da análise gráfica, serão apresentadas e discutidas algumas estatísticas descritivas.

Todos os *outputs*, gráficos e tabelas, que se apresentam foram obtidos recorrendo ao *software* Eviews.

3.1. APRESENTAÇÃO DAS SÉRIES TEMPORAIS

Em termos de aplicabilidade prática, para alcançar os objectivos traçados, os dados em análise foram retirados do DS-Market (*DataStream Market*) e respeitam a índices preços de fecho relativos a mercados accionistas europeus, criados pela Thomson Reuters® (ver Anexo B).

Basicamente, os dados consistem em registos diários (cinco dias/semana) representativos do comportamento dos mercados accionistas de seis países diferentes da zona Euro: Alemanha (TOTMKBD(PI)), Espanha (TOTMKES(PI)), França (TOTMKFR(PI)), Grécia (TOTMKGR(PI)), Itália (TOTMKIT(PI)) e Portugal (TOTMKPT(PI)), cujas series serão denotados, neste trabalho, pelas siglas GER, SP, FR, GR, IT e PT, respectivamente.

Os dados em estudo respeitam ao período temporal de entre 1 de Janeiro de 2002 e 13 de Julho de 2010, num total de 2226 observações por série. Saliente-se, no entanto, que a indexação de preços das séries em estudo teve início em períodos temporais diferentes⁵²: Alemanha, França, e Itália em Janeiro de 1973, Espanha em Janeiro de 1986, Grécia em Janeiro de 1988 e Portugal em Janeiro 1990.

Na análise em questão, com vista minimizar as volatilidades, consideraram-se inicialmente os dados logaritmizados (LGER, LSP, LFR, LGR, LIT e LPT) e,

⁵² Informação disponível em http://thomsonreuters.com/content/financial/pdf/i_and_a/indices/datastream_global_equity_manual.pdf

posteriormente, analisaram-se as diferenças logaritmizadas, eliminando, assim, a não estacionariedade das séries, ou seja, tomamos para cada série as diferenças logaritmizadas (DL, DLGER, DLSP, DLFR, DLGR, DLIT e DLPT) calculadas como $DLd_t = (Ld_t - Ld_{t-1})$, onde, para cada observação, d_t , $Ld_t = \ln(d_t)$.

3.2. ANÁLISE GRÁFICA E ESTATÍSTICAS DESCRITIVAS

No que respeita à análise das séries, começemos por analisar a representação gráfica das mesmas e, posteriormente, observemos algumas características mediante análise de estatísticas descritivas.

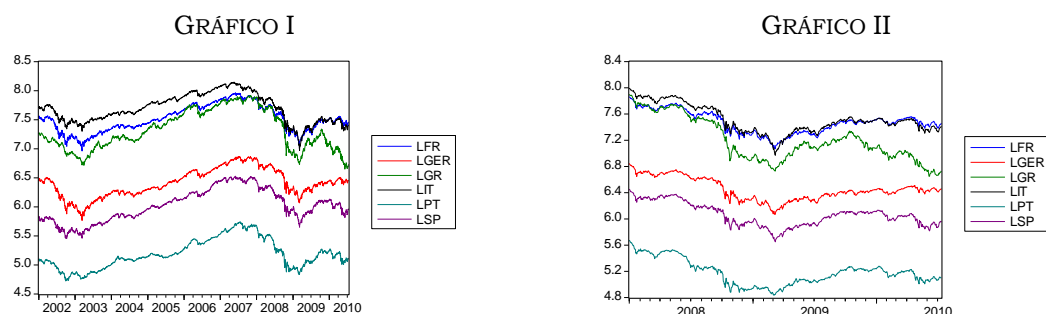


FIGURA 9 – Representação gráfica das séries logaritmizadas

As séries apresentam um padrão não-linear com pelo menos dois regimes distintos. O primeiro regime pode ser considerado até ao final do ano 2007 e o segundo a partir de 2008.

Tendo em conta a Figura 9, em especial o Gráfico I, é possível observar uma queda no ritmo de crescimento entre 2002 e meados de 2003. Se nos recordarmos, tal foi notável nos mercados mundiais, resultante, em parte, da instabilidade política e económica que se sucedeu aos atentados de 11 de Setembro de EUA e a guerra de Afeganistão. A partir de 2004 e até 2007, observa-se uma conjuntura económica favorável que se reflecte numa evolução crescente nos mercados europeus. Já o ano de 2007 salienta uma forte instabilidade nos mercados financeiros devido à Crise do *subprime*⁵³ e a evolução crescente dos preços das matérias-primas. Entre outros factos, verifica-se uma queda na

⁵³ A Crise do *subprime* uma crise financeira desencadeada em 2006, resultante da quebra de instituições de crédito dos Estados Unidos, as quais concediam empréstimos hipotecários de alto risco. Tal situação conduziu a uma situação de insolvência de vários bancos, tendo tido fortes repercussões nas bolsas de valores de todo o mundo. A crise foi revelada ao público a partir de Fevereiro de 2007, como uma crise financeira, constituindo um prenúncio da crise económica de 2008.

generalidade dos mercados accionistas internacionais, e, em particular, europeus, como se observa no gráfico.

A crise no sistema financeiro à escala mundial teve consequências negativas, passando de uma crise financeira localizada a uma crise económica global, com várias economias a entrar em recessão no final de 2008. O quase colapso do sistema financeiro a escala mundial teve consequências bastante negativas, onde grandes instituições financeiras entram em ruptura financeira, como *Bear Sterns* e *Lehman Brothers*.

Depois da queda acentuada de 2008, a instabilidade prolongou-se até ao primeiro trimestre de 2009. Nos trimestres seguintes, verifica-se uma ligeira recuperação, mas ainda com a presença de uma volatilidade acentuada. Em 2009, surge nova instabilidade nos mercados, associada às dívidas soberanas dos países europeus, sobretudo na sequência da crise da dívida pública grega e da redução das classificações de *rating* de vários países europeus, fenómenos que se prolongam para os anos subsequentes.

Assim, face à notável quebra, é possível considerar os dois regimes referidos no início desta descrição. Precisamente, a existência de regimes com diferenças (abruptamente) acentuadas, constituem um entrave à implementação imediata de modelos e técnicas lineares, no contexto da inferência clássica. Deste modo, na análise desenvolvida em secções seguintes, será feito, primeiramente, o estudo inicial para o período global considerado, de 2002 até meados de 2010 (1:2002 – 7:2010), e em seguida para período particular acima identificado como um possível segundo regime, com início em Janeiro de 2008 e termo em Julho de 2010 (1:2008 – 7:2010).

Antes da referida análise, observemos ainda algumas estatísticas relativas às séries em estudo no período global em análise, sendo que em Anexo (Tabela 12 – Anexo C) encontramos dados referentes ao período 1:2008 – 7:2010.

	LFR	LGER	LGR	LIT	LPT	LSP
Mean	7.509774	6.410191	7.318596	7.690352	5.188821	6.019162
Median	7.490876	6.407680	7.224316	7.671594	5.139087	6.002874
Maximum	7.964677	6.868453	7.914117	8.149345	5.743580	6.529521
Minimum	6.961230	5.765411	6.649917	6.970570	4.725439	5.456474
Std. Dev.	0.230538	0.241559	0.339633	0.247236	0.252986	0.277769
Skewness	0.143291	-0.005534	0.166900	0.027129	0.435492	0.131382
Kurtosis	2.094143	2.361120	1.798601	2.082230	2.316125	2.054841
Jarque-Bera	83.72597	37.86891	144.2061	78.39646	113.7391	89.25980
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	16716.76	14269.09	16291.19	17118.72	11550.32	13398.66
Sum Sq. Dev.	118.2542	129.8308	256.6551	136.0047	142.4046	171.6709
Observations	2226	2226	2226	2226	2226	2226

TABELA 5 – Estatísticas descritivas das séries logaritmizadas

Tendo em conta as estatísticas observadas na Tabela 5, é evidente a não normalidade de cada uma das séries (Jarque-Bera), com um *p-value* nulo em cada um dos casos (rejeitando-se a hipótese nula de normalidade para todos os níveis de significância). Verifica-se, ainda, um coeficiente de curtose (*Kurtosis*) baixo (valores menores que 3, logo tratam-se de séries platicúrticas) e alguma assimetria (*Skewness*). Estas mesmas conclusões podem ser validadas graficamente, pela observação do histograma respeitante a cada série (Figura 12 – Anexo C).

No que respeita à correlação entre as séries, é evidente uma forte correlação positiva ao longo do período de tempo considerado, como se depreende pela Tabela 6.

	LFR	LGER	LGR	LIT	LPT	LSP
LFR	1.000000	0.963770	0.930283	0.909351	0.964823	0.948421
LGER	0.963770	1.000000	0.868304	0.788245	0.947866	0.946647
LGR	0.930283	0.868304	1.000000	0.938039	0.918311	0.896283
LIT	0.909351	0.788245	0.938039	1.000000	0.852343	0.799123
LPT	0.964823	0.947866	0.918311	0.852343	1.000000	0.969105
LSP	0.948421	0.946647	0.896283	0.799123	0.969105	1.000000

TABELA 6 – Matriz de Correlação entre as séries logaritmizadas

Com vista a validar as mesmas conclusões para o período 1:2008 – 7:2010, foi feita uma análise idêntica, cujas estatísticas se apresentam em anexo (Tabela 13 – Anexo C), sendo que os resultados obtidos foram idênticos, pelo que as considerações tecidas são igualmente válidas para este período em particular.

No que respeita às séries em diferenças logaritmizadas, apresenta-se de seguida uma análise em toda idêntica à feita para o caso anterior.

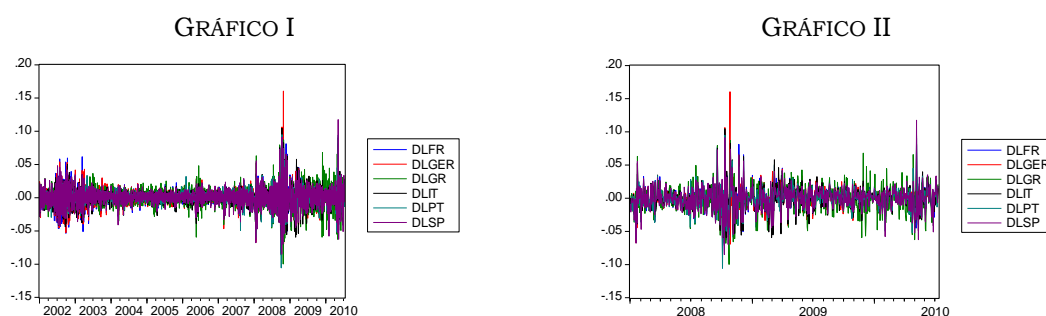


FIGURA 10 – Representação gráfica das séries em DL

Pela mancha gráfica conjunta das seis séries, dada a concentração da mesma, depreende-se o comportamento em todo idêntico em cada uma das séries. Uma análise mais detalhada pode ser feita a partir do gráfico e estatísticas descritivas apresentadas em anexo (Figura 13 – Anexo C), da qual destacamos a não normalidade das mesmas a

avaliar pelo valor do p -value (rejeitamos a hipótese nula de normalidade para todos os níveis de significância), sendo que neste caso o valor referente à curtose é muito superior ao desejável, pelo que as séries são leptocúrticas.

Finalmente, foi feita uma análise à causalidade bivariada de Granger para este conjunto de séries (séries em DL, que são estacionárias, conforme descrito mais à frente), tendo como intuito verificar se existe alguma variável com maior impacto. Face aos valores observados apresentado no Anexo B, é evidente uma bicausalidade entre cada par de variáveis, dado que, em todos os casos, é rejeitada a hipótese nula da não causalidade para um nível de significância de 5%.

3.3. ESTUDO DA ESTACIONARIEDADE

Tal como referido ao longo do Capítulo 2, uma importante característica a estudar nas séries é a estacionariedade, ou não, das mesmas, pelo impacto que tal tem nas etapas subsequentes do estudo.

Deste modo, apesar da inferência feita pela análise gráfica correspondente a cada série, as mesmas conjecturas carecem de uma validação, tendo-se usado, para tal, testes de raiz unitária. Além do Teste ADF (descrito na Secção 2.2.2.), foi feita, também, à análise à existência de raiz unitária usado o Teste de *Phillips-Perron* (PP), apresentando-se, para o efeito, as estatísticas relativas aos dois testes.

Pela análise gráfica (Figura 10) das séries em log-níveis (logaritmizadas) relativas ao período total em estudo, depreendemos que as mesmas parecem ser não estacionárias. Tal facto é validado tendo em conta os Testes ADF e PP (Tabela 7), aceitando-se a hipótese nula de existência de raiz unitária (séries não estacionárias) para qualquer nível de significância.

DADOS RELATIVOS AO PERÍODO DE 2002 A 2010			DADOS RELATIVOS AO PERÍODO DE 2008 A 2010		
Séries	p -value (ADF)	p -value (PP)	Séries	p -value (ADF)	p -value (PP)
LFR	0.5725	0.6399	LFR	0.1522	0.1592
LGER	0.6380	0.6542	LGER	0.1511	0.1550
LGR	0.9083	0.9059	LGR	0.4822	0.4809
LIT	0.7556	0.7565	LIT	0.1920	0.1922
LPT	0.7513	0.7326	LPT	0.0934	0.0922*
LSP	0.6635	0.6963	LSP	0.1799	0.1914

* Rejeita-se H_0 para um nível de significância de 10%

TABELA 7 – Resumo dos resultados relativos ao teste à existência de raiz unitária para as séries logaritmizadas

Já no período de 1:2008 – 7:2010, a análise gráfica pode levantar algumas dúvidas, facto que se confirma tendo em conta a mesma Tabela, cujos valores referentes aos *p-values* são substancialmente mais baixos. Neste caso, apenas podemos aceitar a existência de raiz unitária, para todas as séries, com nível de significância de 5% (pois, para um nível de significância de 10% rejeitávamos a hipótese nula para a série referente a Portugal).

Já quanto às séries em DL, da análise gráfica relativa a ambos os períodos em questão (Figura 10), conjectura-se que as mesmas são estacionárias, facto que é completamente verificado tendo em conta os testes de raiz unitária realizados, cujo valor do *p-value* nos permite rejeitar a hipótese nula para qualquer nível de significância (Tabela 8).

DADOS RELATIVOS AO PERÍODO DE 2002 A 2010			DADOS RELATIVOS AO PERÍODO DE 2008 A 2010		
Séries	<i>p-value</i> (ADF)	<i>p-value</i> (PP)	Séries	<i>p-value</i> (ADF)	<i>p-value</i> (PP)
LFR	0.0001*	0.0001*	LFR	0.0000*	0.0000*
LGER	0.0001*	0.0001*	LGER	0.0000*	0.0000*
LGR	0.0000*	0.0000*	LGR	0.0000*	0.0000*
LIT	0.0000*	0.0001*	LIT	0.0000*	0.0000*
LPT	0.0001*	0.0001*	LPT	0.0000*	0.0000*
LSP	0.0001*	0.0001*	LSP	0.0000*	0.0000*

* Rejeita-se H_0 para um nível de significância de 1%, 5% e 10%

TABELA 8 – Resumo dos resultados relativos ao teste à existência de raiz unitária para as séries em DL

Concluindo, as seis séries temporais em análise são não estacionárias em níveis, mas estacionárias nas primeiras diferenças, ou seja, integradas de primeira ordem. Tal, leva-nos, com certa pertinência, à procura de relações de equilíbrio ao longo prazo entre os mercados analisados no nosso estudo empírico.

4. ESTIMAÇÃO DA ORDEM DE COINTEGRAÇÃO E MODELOS ECONOMÉTRICOS

Feita uma análise gráfica descritiva das variáveis em estudo, passemos a parte prática crucial do nosso trabalho, estimação clássica e bayesiana de um sistema cointegrado VAR e respectiva ordem de cointegração.

Nas páginas seguintes, apresenta-se uma reflexão sobre os resultados obtidos referentes à estimação da ordem de cointegração e consequente estimação de modelos VAR/VECM ou BVAR/BVECM, em função da metodologia usada (clássica ou bayesiana, respectivamente), cujos respectivos outputs constam nos Anexos D e E.

4.1. ESTIMAÇÃO/INFERÊNCIAS USANDO A METODOLOGIA CLÁSSICA

No que respeita a abordagem clássica, todo o processo de estudo, estimação da ordem de cointegração e modelos, foi desenvolvido recorrendo às potencialidades do *software Eviews*, cujos *outputs* se apresentam todos no Anexo D.

Seguindo um procedimento idêntico ao da secção anterior, pelas razões já justificadas, sucedendo à análise para o período global em estudo (1:2002 – 7:2010), far-se-á uma análise para o período correspondente aos dois regimes considerados: 1:2002 – 1:2008 e 1:2008 – 7:2010.

O primeiro passo na estimação do modelo VAR, para séries estacionárias, consiste na determinação do número de *lags* óptimo a considerar para, em seguida, se proceder à estimação da ordem de cointegração, segundo Johansen. Neste processo, procurando fazer um estudo mais completo, com vista à obtenção de conclusões mais assertivas, foram considerados dois casos, usando critérios de informação distintos: SC e AIC.

Na Tabela seguinte apresenta-se uma síntese da informação essencial constante nas Tabelas 15 – 18 e 21 – 27, do Anexo D, referente aos dois períodos em causa.

CRITÉRIOS DADOS (PERÍODO)	CRITÉRIO: SC			CRITÉRIO: AIC		
	N.º de <i>lags</i>	N.º de vectores cointegrantes ⁵⁴		N.º de <i>lags</i>	N.º de vectores cointegrantes ⁵⁶	
		Traço	Max.		Traço	Max.
1:2002 – 7:2010	1	1	1	20	0	0
1:2002 – 1:2008	1	1	1	3	1	0
1:2008 – 7:2010	1	0	1	8	0	0

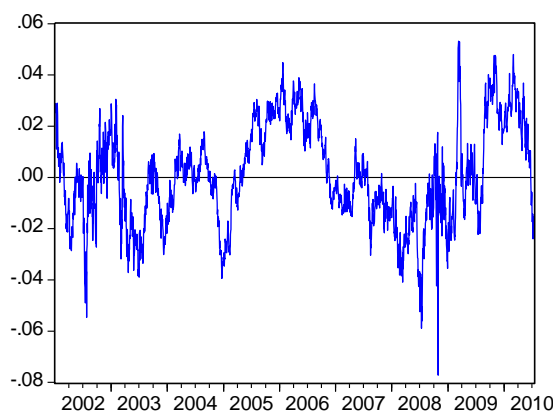
TABELA 9 – Resumo dos resultados relativos à estimação do número de *lags* óptimo e ordem de cointegração (abordagem clássica)

Da análise da tabela, verificamos que as estatísticas obtidas com base nos dois critérios não são consensuais.

No período global (1:2002 – 7:2010), pelo SC, cujo número de *lags* é dado por defeito (1 *lag*), a metodologia clássica de Johansen sugere a existência de um vector cointegrante, ao passo que, considerando AIC, cujo número de *lags* é dado por excesso (20 *lags*), não existe relação ao longo prazo entre as séries.

Já para o período 1:2002 – 1:2008, concordante com o SC, segundo a estatística do traço para o AIC, existe uma relação de cointegração. O mesmo não acontece para o segundo período em análise, onde parece não existir relação de longo prazo, pois, considerando o SC (segundo o teste do valor próprio Max), existe um vector cointegrante. Porém, recorde-se que segundo a bibliografia, o teste do traço é mais poderoso.

Em função dos dados obtidos e das relações de cointegração consideradas, foi feita uma estimação de modelos VAR/VECM, consoante a existência, ou não, de relações de cointegração entre as séries.



— 1*LFR-0.519358*LGER+0.120920*LGR-0.448852*LIT-0.034611*LPT-0.137989*LSP-0.602836

FIGURA 11 – Representação gráfica da relação de longo prazo (1:2002 – 7:2010)

⁵⁴ Considerando a opção *intecept* e *no trend*.

Considerando a relação de cointegração sugerida pelo SC no período global, 1:2002 – 7:2010, representada na Figura 11, verificamos que o gráfico oscila bastante em torno do equilíbrio a longo prazo, sendo evidente a quebra registada no ano 2008 comparativamente à linha de equilíbrio, decorrente da crise financeira registada.

Em cada caso foi feita uma estimativa do modelo em questão⁵⁵, cujos gráficos relativos à estabilidade se encontram em anexo (Anexo D). Para os períodos 1:2002 – 7:2010 e 1:2008 – 7:2010, os modelos que evidenciam uma maior estabilidade são os modelos VAR (considerando que as séries não são estacionárias e nem cointegradas), que representamos, por VAR(20,6) – 20 *lags* e 6 variáveis – e VAR(8,6) – 8 *lags* e 6 variáveis –, respectivamente, uma vez que as raízes inversas do polinómio característicos estão contidas no interior do círculo unitário (Figura 15 e Figura 18).

Para terminar, foi ainda analisada a causalidade de Granger, procurando compreender a relação de causalidade entre as variáveis. Para os modelos VAR considerados anteriormente, verificamos que no caso do VAR(20,6), para o período de 1:2002 – 7:2010, de um modo geral os valores relativos aos *p-values* são baixos, pelo que rejeitamos a hipótese nula de exclusão do modelo para qualquer nível de significância (Tabela 20), ou seja, de um modo geral, todas as variáveis têm um contributo na explicação das outras (interdependência dos mercados). Pensado nos últimos anos, aí é verificado um comportamento diferente, sendo que a França e a Alemanha exercem uma forte dependência sobre os outros mercados bolsistas (facto a confirmar a conjuntura actual), uma vez que rejeitamos a hipótese nula de exclusão do modelo, considerando cada variável (Tabela 29).

4.2. ESTIMAÇÃO/INFERÊNCIAS USANDO A METODOLOGIA BAYESIANA

A estimação da ordem de cointegração, segundo uma abordagem bayesiana, foi feita usando códigos de programação em *MatLab*, desenvolvidos por Warne (2006), resultante do trabalho desenvolvido por Villani (2005) e pelo próprio, cujos *outputs* se apresentam no Anexo E. É possível observar nos mesmos algumas considerações e explicações

⁵⁵ É importante referir que, neste trabalho, não foi considerada a análise ao ajustamento dos modelos estimados. Tal tornaria bastante extensa a análise e sai fora do âmbito central do estudo. No entanto, sublinhe-se que, para verificar e validar a adequabilidade dos modelos, carece ser feita uma análise completa aos resíduos, tendo em conta os vários pressupostos a verificar (sintetizados em Mendes (2010)).

relativas ao processo de computação e aspectos envolvidos, nomeadamente sobre as distribuições *a priori* e consequente *posteriori*.

À semelhança com a apresentação feita para o caso clássico, apresenta-se na tabela seguinte uma síntese das principais conclusões constantes nas Tabelas 30 – 33.

CRITÉRIOS DADOS (PERÍODO)	CRITÉRIO: SC			CRITÉRIO: AIC		
	N.º de <i>lags</i>	N.º de vectores cointegrantes		N.º de <i>lags</i>	N.º de vectores cointegrantes	
		FML	ML		FML	ML
1:2002 – 7:2010	2	0	0	19	0	0
1:2008 – 7:2010	1	0	1	8	0	0

TABELA 10 – Resumo dos resultados relativos à estimação do número de *lags* óptimo e ordem de cointegração (abordagem bayesiana)

Da análise da tabela verificamos que os resultados são bastante mais consensuais. Apenas, para o período 1:2008 – 7:2010, existe uma divergência quando a computação foi baseada na *Marginal Likelihood* (ML), considerando o SC, estimando a possibilidade de existência de um vector cointegrante.

No entanto, embora possa existir um vector cointegrante no período de tempo parcial analisado, no que respeita ao período global não existe nenhuma relação de cointegração.

No que respeita à estimação dos modelos BVAR/BECM, os mesmos foram obtidos utilizando uma distribuição *a priori* específica, conhecida na literatura como *priori de Minnesota*⁵⁶ e cuja *a posteriori* foi obtida via algoritmo de *Gibbs*. Os procedimentos práticos e códigos implementados em *MatLab*, foram os descritos em LaSage (1999).

Os *outputs* gráficos relativos aos modelos estimados constam no Anexo E (Figura 19 e Figura 20). Com o objectivo de experimentação apresentam-se somente os valores reais e estimados de três países, os dois mercados com maior impacto (Alemanha e França, conforme o visto na Secção 3.2) e o caso de Portugal. Da análise dos mesmos, verificamos que as previsões obtidas nos dois modelos são relativamente assertivas, pois as linhas correspondentes ao actual e ao predito estão bastante próximas. Em partícula, se tivermos em conta o modelo BVAR estimado (Figura 20), verificamos que, além de traçados bastante próximos entre o actual e o predito, a representação gráfica correspondente aos erros é bastante razoável, com uma ordem de grandeza reduzida⁵⁷.

⁵⁶ Ver Doan, Litterman e Sims (1984).

⁵⁷ A validação do modelo, carecia de uma análise detalhada aos erros.

4.3. ANÁLISE COMPARATIVA DOS RESULTADOS OBTIDOS

Na literatura científica, vários são os estudos que utilizam as técnicas de cointegração para analisar as relações de longo prazo entre índices dos preços dos mercados internacionais, dadas as implicações determinantes que a cointegração tem na avaliação de activos financeiros (*asset pricing*). Por um lado, a existência de cointegração entre vários índices de preços significa que os mercados partilham uma tendência estocástica comum. Por outro lado, a cointegração entre índices conduz a rejeição da hipótese de mercados eficientes⁵⁸, pois, a cointegração induz predictabilidade ao curto prazo dos preços via o mecanismo correcção do erro.

Deste modo, os estudos empíricos mais recentes sobre a análise dos co-movimentos de séries temporais, sustentam cada vez mais a tese de que os índices dos preços dos mercados financeiros internacionais não são cointegrados. Alguns dos trabalhos desenvolvidos onde podemos constatar tal facto são: Huang e Fok (2001), Pascual (2003), Davies (2006) e Li (2006).

Uma razão importante para explicar os resultados empíricos em questão reside no facto das técnicas lineares usuais poderem revelar-se inadequadas na presença de dinâmicas não *standard*, não linearidades ou mesmo quebras estruturais (situação esta que parece caracterizar as séries em estudo). Nos artigos respectivos, Li (2006) utiliza testes não-lineares de cointegração e Davies (2006) aplica cointegração com mudanças de regime. Ambos os autores sugerem que, de facto, é possível existir cointegração temporal, por vários períodos de tempo, mas não se pode garantir a convergência para um equilíbrio único ao longo do tempo total. Apesar disto, vários são os trabalhos, principalmente os que são desenvolvidos com base em metodologias clássicas) onde se encontram evidências de correlação serial muito forte ao curto prazo e relações causais entre os índices.

Estes factos vêm clarificar algumas questões anteriormente levantadas, sobretudo no que respeita ao tema central – determinação da ordem de cointegração – uma vez que, existem algumas divergências nos resultados obtidos pelas duas metodologias.

Procurado sintetizar a informação, vejamos o quadro seguinte.

⁵⁸ A hipótese dos mercados eficientes baseia-se na crença de que as cotações dos activos cotados (acções, obrigações, ...) reflectem toda a informação conhecida. Com efeito, esta hipótese teoriza que é impossível bater consistentemente os mercados usando qualquer informação já disponível no presente, excepto pela sorte.

CRITÉRIOS DADOS (PERÍODO)	METODOLOGIA CLÁSSICA		METODOLOGIA BAYESIANA	
	N.º de <i>lags</i>	N.º de vectores cointegrantes	N.º de <i>lags</i>	N.º de vectores cointegrantes
1:2002 – 7:2010	1 a 20	0 a 1	2 a 19	0
1:2008 – 7:2010	1 a 8	0 a 1	1 a 8	0 a 1

TABELA 11 – Resumo geral dos resultados relativos à estimação do número de *lags* óptimo e ordem de cointegração

Tendo em conta os resultados obtidos, podemos inferir que a metodologia clássica fornece resultados concordantes com a metodologia bayesiana, se considerarmos um período temporal sem quebras estruturais significativas. Isto porque, no período de 2008 a 2010, ambas as metodologias, consoante o número de desfasamentos usados, referem a eventual possibilidade de existência de uma relação cointegrante, se bem que, considerando um número maior de desfasamentos a considerar no modelo, tal hipótese é excluída.

Porém, quando consideramos um período temporal significativo, cujos dados apresentam volatilidades consideráveis, dando indícios de existência de regimes distintos, a metodologia clássica apresenta resultados completamente contraditórios (consoante o número de desfasamentos a incluir no modelo). Num dos casos, constata-se a existência clara de uma relação cointegrante. Já os resultados observados mediante a aplicação dos pressupostos bayesianos são concordantes, concluindo-se a não existência de convergência a longo prazo entre os vários índices.

Assim, por meio de testes econométricos, envolvendo inferência/estimação clássica e bayesiana, mostramos que, apesar de ocorrer alguma (hipotética e pontual) convergência num período temporal, a inferência bayesiana evidencia que os processos são independentes na globalidade, indo, assim, ao encontro dos resultados empíricos que persistem na literatura científica.

CONCLUSÃO

Face a uma motivação inicial para a compreensão da aplicabilidade da abordagem bayesiana na inferência estatística e o particular interesse pela área da Econometria, o objectivo central deste trabalho incidia na comparação de resultados obtidos pela estimação clássica e bayesiana de um sistema cointegrado VAR, com especial ênfase para a determinação da respectiva ordem de cointegração.

Da implementação computacional desenvolvida, onde se consideraram as seis séries temporais correspondentes a índices bolsistas dos seis países do Mercado Europeu da zona Euro escolhido para o efeito (Alemanha, Espanha, França, Grécia, Itália e Portugal), podemos inferir que as duas metodologias são concordantes, se considerarmos um período temporal sem quebras estruturais significativas. Isto porque, no período de 2008 a 2010, considerando um número maior de desfasamentos a incluir no modelo, ambas referem a inexistência de uma relação ao longo prazo entre as variáveis. Contudo, reduzindo o número de desfasamentos, é possível encontrar uma relação de cointegração.

Tal concordância deixa de ser verificada quando aumentamos a dimensão da amostra, passando a considerar um período temporal onde dados apresentam volatilidades consideráveis, dando indícios da existência de regimes distintos. Neste caso, a metodologia clássica apresenta resultados completamente contraditórios (consoante o número de desfasamentos a incluir no modelo). Num dos casos, constata-se a existência clara de uma relação cointegrante, o que é contraditório quando tida em conta a literatura científica sobre o tema, que sustenta cada vez mais a tese de que os índices dos preços dos mercados financeiros internacionais não são cointegrados. Já os resultados observados, mediante a aplicação dos pressupostos bayesianos, são concordantes na não existência de convergência a longo prazo entre os vários índices.

Assim, por meio de testes econométricos, envolvendo inferência/estimação clássica e bayesiana, mostramos que, apesar de ocorrer alguma (hipotética e pontual) convergência num período temporal, a inferência bayesiana evidencia que os processos são independentes na globalidade, indo, assim, ao encontro dos resultados empíricos que persistem na literatura científica.

No que respeita aos modelos econométricos em si, partindo da não existência de relações cointegrantes, os modelos a considerar são os modelos VAR ou BVAR,

consoante o tipo de abordagem seguida. De uma análise dos resultados obtidos, verificamos que, nos últimos anos, a França e a Alemanha exercem efectivamente uma forte dependência sobre os outros mercados bolsistas uma vez que rejeitamos a hipótese nula de exclusão do modelo, considerando cada variável, quando analisada a causalidade entre as variáveis em estudo.

Pelos requisitos necessários à implementação prática que nos permitiu chegar aos resultados referidos, pode dizer-se que o trabalho desenvolvido constitui uma introdução ao estudo e aplicação das técnicas bayesiana de inferência estatística e de simulação MCMC.

Esta última parte é talvez uma das limitações a apontar ao trabalho desenvolvido. Pelo facto da parte computacional ser um domínio amplo e complexo de estudo (algo que por si só seria um tema amplo de trabalho), tal carecia de um maior investimento, no sentido de implementar computacionalmente desenvolvimentos teóricos com vista à obtenção de uma informação mais completa dos modelos econométricos estimados e consequente análise e validação, ou não, dos mesmos. Como se referiu, tal não foi feito pela extensão que o estudo em si comportava, pelo que será, seguramente, um caminho de continuidade da pesquisa desenvolvida, uma estimação e estudo mais detalhado e completo aos modelos VAR e BVAR (possivelmente considerando ou não a sua extensão aos modelos que incorporam a modelação explícita de relações a longo prazo).

Com efeito, além de uma validação dos modelos clássicos obtidos, através da análise da componente residual, propõe-se um investimento na obtenção de uma informação mais completa relativa ao modelo bayesianos e consequente estudo/validação dos mesmos, finalizando com um confronto à adequabilidade dos mesmos face ao observável e consequente desempenho em termos de previsão (algo que é sempre interessante analisar).

Além desta questão emergente, dado que o conjunto de dados apresentava quebras estruturais, no seguimento do que em parte foi feito, outros aspectos seriam interessantes analisar.

Por um lado, face a algumas das limitações conhecidas da metodologia clássica, como a dificuldade em fazer inferências assertivas com base num conjunto de dados muito reduzido e a dificuldade em lidar com situações de sobreparametrização, seria interessante analisar com maior detalhe estes dois casos extremos. Assim, partindo do conjunto global de dados, procurar-se-ia tirar conclusões mais exactas sobre a forma como os modelos bayesianos conseguem, ou não, dar resposta aos problemas referidos para o conjunto de observações consideradas em análise.

Uma outra abordagem a analisar, face a natureza não-linear dos dados e no seguimento do referido na literatura (Davis, 2006), seria um estudo baseado nos modelos *Markov-Switching* para a amostra considerada.

Em suma, reconhecendo o trabalho já desenvolvido, nomeadamente o estudo relativo à estimação da ordem de cointegração, identificamos nele vários pontos de partida para novos estudos, passíveis de serem desenvolvidos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ----- (2004). *EViews 5 User's Guide*. Quantitative Micro Software, LLC.
- [2] Bauwens, L., & Lubrano, M. (1996). *Identification restrictions and posterior densities in cointegrated Gaussian VAR systems*. *Advances in Econometrics* 11, Part B, (JAI Press, Greenwich) 3-28.
- [3] Box, G., & Jenkins, G. (1976). *Time Series Analysis: forecasting and control*. San Francisco: Holden-Day.
- [4] Costa, W. (2004). *Técnicas Bayesianas para Engenharia Elétrica*. Obtido em 18 de Abril de 2011, de <http://www.cpdee.ufmg.br/~wadaed/Pesquisa/TecBayesianas.pdf>
- [5] Davies, A. (2006). *Testing for international equity market integration using regime switching cointegration techniques*. *Review of Financial Economics* 15, 305—321.
- [6] DeJong, D. (1992). *Co-integration and trend-stationarity in macroeconomic time series*. *Journal of Econometrics*, 52, 347-370.
- [7] Doan, T., Litterman, R., & Sims, C. (1984). *Forecasting and conditional projections using a realistic prior distribution*. *Econometric Reviews* 3: 1–100.
- [8] Ehlers, R. (2007). *Inferência Bayesiana*. Obtido em 10 de 7 de 2011, de <http://www.icmc.usp.br/~ehlers/bayes/bayes.pdf>
- [9] Engle, R., & Granger, C. (1987). *Co-integration and error correction: representation, estimation and testing*. *Econometrica* 55, 251-76.
- [10] Esquível, M. (2009). *Teoria Elementar das Cadeias de Markov*. Obtido em 22 de Maio de 2011, de <http://ferrari.dmat.fct.unl.pt/personal/mle/DocPE/PE0809/Docmts-3/CadeiasdeMarkov.pdf>
- [11] Feldman, R., & Valdez-Flores, C. (2010). *Applied Probability and Stochastic Processes* (2^a ed.). Berlin: Springer.
- [12] Felix, R., & Nunes, L. (2002). *Forecasting euro area aggregates with Bayesian VAR and VECM models*. Working Paper 4-03, Banco de Portugal.
- [13] Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.

- [14] Gamerman, D., & Lopes, H. (2006). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2^a ed.). London: Chapman and Hall.
- [15] Gelfand, A., & Smith, A. (1990). *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 85, 972-985.
- [16] Geweke, J. (1989). *Bayesian inference in econometric models using Monte Carlo integration*. Econometrica, 57, 1317-1339.
- [17] Geweke, J. (1996). *Bayesian reduced rank regression in econometrics*. Journal of Econometrics 75, 121-146.
- [18] Glynn, P., & Asmussen, S. (2007). *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.
- [19] Grimmett, G., & Stirzaker, D. (1992). *Probability and Random Processes* (3^a ed.). New York: Oxford University Press.
- [20] Hoel, P., Port, S., & Stone, C. (1987). *Introduction to Stochastic Processes*. Waveland Press.
- [21] Huang, B.-N., & Fok, C. (2001). *Stockmarket integration: an application of the stochastic permanent breaks model*. Applied Economics Letters 8, 725-729.
- [22] Jaramillo, P. (2008). *Estimación de VAR Bayesianos para la Economía Chilena*. Working Paper 508, Banco Central do Chile.
- [23] Johansen, S. (1988). *Statistical Analysis of Cointegration Vectors*. Journal of Economic Dynamics and Control.
- [24] Johansen, S., & Juselius, K. (1990). *Maximum Likelihood Estimation and Inference on Cointegration — with Applications to the Demand for Money*. Oxford Bulletin of Economics and Statistics, 52, 169–210.
- [25] Johnston, J., & Dinardo, J. (2001). *Métodos Econométricos* (4^a ed.). (M. Hill, F. Ferrão, & R. Menezes, Trads.) McGraw-Hill.
- [26] Kirchgässner, G., & Wolters, J. (2007). *Introduction to Modern Time Series Analysis*. Berlin: Springer.
- [27] Kleibergen, F., & van Dijk, H. (1994). *On the shape of the likelihood/posterior in cointegration models*. Econometric Theory 10, 514-551.
- [28] Koop, G. (1994). *An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates*. Journal of Empirical Finance, 1, 343-364.
- [29] Koop, G. (2004). *Bayesian Econometrics*. London: John Wiley.

- [30] Koop, G. (1991). *Cointegration tests in present value relationships: A Bayesian look at the bivariate properties of stock prices and dividends*. Journal of Econometrics, 49, 105-140.
- [31] Lahiri, K., & Martin, G. (2010). *Bayesian forecasting in economics*. International Journal of Forecasting, 26, 211-215.
- [32] Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing.
- [33] LaSage, J. (1999). *Applied Econometrics using MATLAB*. Obtido em 25 de 7 de 2011, de <http://www.spatial-econometrics.com/html/mbook.pdf>
- [34] Li, X.-M. (2006). *A revisit of international stock market linkages: new evidence from rank tests for nonlinear cointegration*. Scottish Journal of Political Economy, Vol. 53, No. 2.
- [35] Mendes, D. (2010). *Econometria dos Mercados Financeiros, Apontamentos das aulas*. Mestrado de Matematica Financeira, ISCTE.
- [36] Murteira, B. (1990). *Probabilidades e Estatística* (Vol. I e II). Mc Graw-Hill.
- [37] Nascimento, M. (2009). *O uso de simulação de Monte Carlo via Cadeias de Markov no melhoramento genético*. Minas Gerais: Tese de Mestrado, Universidade Federal de Viçosa.
- [38] Pascual, A. (2003). *Assessing European stock markets (co)integration*. Economics Letters 78, 197–203.
- [39] Paulino, C., Turkman, M., & Murteira, B. (2003). *Estatística Bayesiana*. Lisboa: Fundação Calouste Glubenkian.
- [40] Shonkwiler, R., & Mendivil, F. (2009). *Explorations in Monte Carlo Methods*. New York: Springer.
- [41] Tierney, L. (1994). *Markov Chain for Exploring Posterior Distributions*. The Annals of Statistics, 22, 1701-1762.
- [42] Turkman, M. (s.d.). *How and when does the Gibbs Sampler Work*. Obtido em 15 de 4 de 2011, de www.ine.pt
- [43] van Dijk, H., Kloek, T., Heij, C., Franses, P., & de Boer, P. (2004). *Econometric Methods with Applications in Business and*. Oxford University Press.
- [44] Villani, M. (2005). *Bayesian Reference Analysis of Cointegration*. Econometric Theory, 21, 326–357.
- [45] Warne, A. (2006). *Bayesian inference in cointegrated VAR models*. Working Paper 692, European Central Bank.

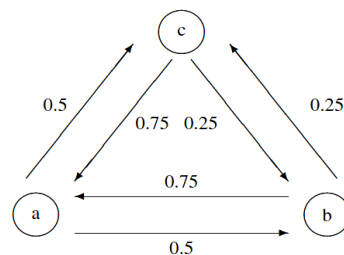
ANEXO A

A.1. EXEMPLOS RELATIVOS A CADEIAS DE MARKOV EM TEMPO DISCRETO

EXEMPLO 1.1.: Cadeia de Markov (um exemplo em contexto real) ⁵⁹

Um vendedor mora na cidade a e dá assistência comercial às cidades a, b e c , mudando de cidade semana a semana. Assim, após uma semana numa dada cidade, na semana seguinte segue, obrigatoriamente, para uma cidade diferente. Quando está na sua cidade (cidade a), não existe nenhuma condicionante sobre qual a cidade a visitar em seguida. Com tal, lança uma moeda ao ar, se sair face nacional, vai para b ; se sair face comum segue para a cidade c . Porém, quando se encontra numa das outras duas cidades e pretende definir qual a próxima cidade a visitar, a experiência aleatória realizada é diferente, dada a preferência em ir para casa. Desta forma, quando está nas cidades b ou c , lança simultaneamente duas moedas, se sair face nacional em ambas as moedas, ele vai para a outra cidade (c ou b); em qualquer outro caso, vai para a sua cidade. Concretamente, tendo em conta a teoria acima apresentada, as cidades sucessivas que ele visita formam uma cadeia de Markov com espaço de estados $S = \{a, b, c\}$, onde a variável aleatória X_n é igual a, b ou c de acordo com a cidade onde se encontra na semana n . O diagrama de estado para este sistema é dado pela figura seguinte, cuja matriz de transição, \mathcal{P} , é dada por:

$$\mathcal{P} = \begin{array}{ccc} \downarrow a & \downarrow b & \downarrow c \\ \begin{bmatrix} 0 & 0,5 & 0,5 \\ 0,75 & 0 & 0,25 \\ 0,75 & 0,25 & 0 \end{bmatrix} & \begin{matrix} \leftarrow a \\ \leftarrow b \\ \leftarrow c \end{matrix} \end{array}$$



Repare-se que a diagonal principal da matriz é nula, pois, como é referido no início do enunciado, após uma semana ele é obrigado a seguir para uma outra cidade.

⁵⁹ Adaptado de Feldman e Valdez-Flores (2010)

EXEMPLO 1.2.: Cadeia de Markov Irredutível e não Irredutível ⁶⁰

Seja $X = \{X_n; n = 0, 1, \dots\}$ uma cadeia de Markov com espaço de estados $S = \{1, 2, 3, 4\}$, cuja matriz de transição e diagrama de estados são dados por:

$$P = \begin{array}{cccc|l} \downarrow 1 & \downarrow 2 & \downarrow 3 & \downarrow 4 & \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0,3 & 0,7 & 0 \\ 0 & 0,5 & 0,5 & 0 \\ 0,2 & 0 & 0,1 & 0,7 \end{array} \right] & \leftarrow 1 \\ & & & & \leftarrow 2 \\ & & & & \leftarrow 3 \\ & & & & \leftarrow 4 \end{array}$$



Neste exemplo, a cadeia é estruturalmente diferente do Exemplo 1.1. Por exemplo, podemos iniciar no estado 4, depois seguir para o estado 3 e não chegar ao estado 1, ou mesmo, iniciar no estado 4 e ir para o estado 1 e ficar sempre lá, não seguindo para nenhum outro estado. Já o Exemplo 1.1 envolve transições onde é sempre possível atingir um novo estado através de um dado estado, isto é, dado um estado i é sempre possível passar para um estado j (com $i \neq j$).

EXEMPLO 1.3.: Convergência da Cadeia ⁶¹

Considerem-se duas urnas, A e B , e três bolas numeradas de 1 a 3, estando, num instante inicial, distribuídas aleatoriamente pelas duas urnas. Considere-se a experiência que consiste em seleccionar um número do conjunto $\{1, 2, 3\}$ e em seguida remover a bola identificada com esse número da urna respectiva. Em seguida, selecciona-se, aleatoriamente, uma das urnas e coloca-se a bola removida dentro de urna seleccionada. Este procedimento é repetido infinitamente, onde as selecções são feitas de forma independente. Seja X_n o número de bolas na urna A no instante n .

A sequência $(X_n)_{n \geq 0}$ é uma cadeia de Markov homogénea em $S = \{0, 1, 2, 3\}$, com função de transição definida pela matriz de transição:

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 \\ 0 & 1/3 & 1/2 & 1/6 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

⁶⁰ Adaptado de Feldman e Valdez-Flores (2010)

⁶¹ Adaptado de Nascimento (2009).

Tendo em conta as definições 1.10 e 1.11 é possível mostrar que a cadeia em causa é irreduzível e aperiódica. Além disso, tendo em conta que $\pi \mathcal{P} = \pi$, podemos concluir que a função estacionária, π , para a cadeia $(X_n)_{n \geq 0}$ é dada por:

$$\pi = (1/8; 3/8; 2/8; 1/8)$$

Temos:

1) Se $\pi^{(0)} = \pi$, então $\pi^{(n)} = \pi$ para todo o $n > 0$, pois $\pi = \pi \mathcal{P}$ e $\pi^{(n)} = \pi^{(n-1)} \mathcal{P}$ para todo $n > 0$.

2) Porém, caso $\pi^{(0)} \neq \pi$, então é possível verificar $\pi^{(n)} \xrightarrow[n]{} \pi$.

Por exemplo, caso $\pi = (1/4; 1/4; 1/4; 1/4)$

- $\pi^{(1)} = (0,167; 0,333; 0,333; 0,167)$
- $\pi^{(3)} = (0,130; 0,370; 0,370; 0,130)$
- $\pi^{(30)} = (0,125; 0,375; 0,375; 0,125)$

Observa-se ainda que $\lim_{n \rightarrow \infty} p_{ij}^n = \pi_j$.

Por exemplo, a matriz de transição com $n = 30$ vem dada por:

$$\mathcal{P}^{(31)} = \begin{bmatrix} 0,125 & 0,375 & 0,375 & 0,125 \\ 0,125 & 0,375 & 0,375 & 0,125 \\ 0,125 & 0,375 & 0,375 & 0,125 \\ 0,125 & 0,375 & 0,375 & 0,125 \end{bmatrix}$$

Pelo que, aumentando significativamente a ordem n , teríamos

$$\|\pi^{(n)} - \pi\| \xrightarrow[n]{} 0$$

EXEMPLO 1.4.: Reversibilidade e estacionaridade de uma distribuição ⁶²

Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov homogénea em $S = \{0, 1, 2\}$ com função de transição:

$$\mathcal{P} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/6 & 1/6 & 1/3 \\ 0 & 3/4 & 1/4 \end{bmatrix}$$

Usando a Definição 1.17, obtém-se:

$$\begin{aligned} \pi_0 \frac{1}{2} &= \pi_1 \frac{1}{6} \\ \pi_1 \frac{1}{3} &= \pi_2 \frac{3}{4} \end{aligned}$$

Logo $\pi_1 = 3\pi_0$ e $\pi_2 = (4/3)\pi_0$. Como as componentes de π devem somar 1 para que ela seja uma distribuição de probabilidade em S , tem-se que:

$$\pi = \left(\pi_0; 3\pi_0; \frac{4}{3}\pi_0 \right) = (3/16; 9/16; 4/16)$$

Logo π é uma distribuição reversível para $(X_n)_{n \geq 0}$. Pode ver-se que π satisfaz a equação $\pi = \pi\mathcal{P}$ e, portanto, é também uma distribuição estacionária para $(X_n)_{n \geq 0}$.

⁶² Adaptado de Nascimento (2009).

A.2. EXEMPLOS RELATIVOS AOS MCMC ⁶³

EXEMPLO 1.5.: Aplicação do algoritmo de *Metropolis-Hastings*

Considere uma variável aleatória discreta $X \sim \pi(1/3; 1/2; 1/6)$.

Estimemos $E(X^2)$ via algoritmo de *Metropolis-Hastings*.

Antes de aplicar o algoritmo de *Metropolis-Hastings* a fim de obter amostras de π , vejamos, explicitamente, todas as componentes que integram o algoritmo, ou seja, $q(x|y)$, $\alpha(x, y)$ e $p(x|y)$ da cadeia do algoritmo.

Sendo $S = \{1, 2, 3\}$ e q escolhida como

$$q(x|y) = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 4/9 & 4/9 & 1/9 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Usando o facto que $\alpha(x, y) = \left(\frac{\pi(y)q(x,y)}{\pi(x)q(x,y)}, 1 \right)$, obtém-se

$$\alpha(x, y) = \begin{bmatrix} 1 & 1 & 2/3 \\ 3/4 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Vindo a matriz de transição (do algoritmo) dada por:

$$\mathcal{P} = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 1/3 & 5/9 & 1/9 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Observa-se que $\pi = \pi\mathcal{P}$, ou seja, π é de facto a única distribuição estacionária da cadeia.

Sendo a estimação de I dada por:

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n X_i^2,$$

podemos verificar a proximidade entre o valor da estimativa obtida pelo algoritmo de *Metropolis-Hastings*, 3,82. E o valor exacto de I , $23/6 \approx 3,83$.

⁶³ Adaptados de Nascimento (2009).

EXEMPLO 1.6.: Aplicação do método de amostragem de Gibbs

Seja $X = (X_1, X_2)$ um vector bidimensional discreto, com distribuição de probabilidade conjunta $\pi(x)$ dada por

$$\pi(x_1, x_2) = \begin{bmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{bmatrix}$$

Obter uma estimativa para o valor esperado da função f definida por

$$h(X_1, X_2) = \frac{1}{1 + e^{-(x_1 + x_2)}}$$

ou seja, estimar um valor para a quantidade

$$I = E_{\pi}[h(X_1, X_2)] = \sum_{x \in A} \pi(x_1, x_2) \frac{1}{1 + e^{-(x_1 + x_2)}},$$

Para que seja possível o uso do algoritmo de amostragem de Gibbs, é necessário determinar as distribuições condicionais completas de X . Desta forma:

$$\pi_1(x_1 | X_2 = x_2) = \begin{cases} (2/3, 1/3) & \text{se } x_2 = 0 \\ (1/3, 2/3) & \text{se } x_2 = 1 \end{cases}$$

e

$$\pi_2(x_2 | X_1 = x_1) = \begin{cases} (2/3, 1/3) & \text{se } x_1 = 0 \\ (1/3, 2/3) & \text{se } x_1 = 1 \end{cases}$$

Assim obtém-se todos os ingredientes necessários para implementação do algoritmo. O valor de I é 0,704 e a estimativa obtida pelo método foi $\hat{I} = 0,706$.

ANEXO B ⁶⁴

A informação disponibilizada pela Thompson Reuters[®] (TR) é utilizada pelos profissionais da área financeira em todo mundo por ser considerada informação fidedigna.

Um dos serviços fornecidos pela TR está relacionado com a disponibilização de índices, com fonte no *DataStream*. Para a construção do índice de capitalização do mercado, são recolhidos dados dos títulos negociados em cada mercado (bolsa de cada país), sendo posteriormente aplicada uma metodologia própria da TR.

No que respeita a essa metodologia, antes de ser construído o índice é aplicado um teste de liquidez de forma a identificar quais os activos elegíveis. Identificados esses activos, é aplicada a regra de capitalização de mercados, a qual estabelece que o índice de cada mercado deve ser composto por, aproximadamente, 97,68% do valor total em bolsa. Dessa forma são excluídos os títulos referentes a micro capitalizações.

Quanto à formulação dos preços, o índice de capitalização de mercado usa o valor de componente (preço \times acções *free float*) como um factor de ponderação.

Matematicamente, o valor do preço do índice, PI , em cada momento t , é dado por:

$$PI_t = 100 \times \frac{\sum_{i=1}^n (p_{i,t} \cdot q_{i,t} \cdot r_{i,t})}{\sum_{i=1}^n (p_{i,0} \cdot q_{i,0} \cdot r_{i,0})}$$

onde:

- n denota o número de activos consideradas no índice;
- $p_{i,t}$ denota o preço do i -ésimo activo, $i = 1, 2, \dots, n$, no momento $t = 0, 1, \dots, T$ (com $t = 0$ a denotar o momento inicial);
- $q_{i,t}$ denota o ajustamento para o activo no momento t ;
- $r_{i,t}$ é igual a 1, caso se trate do índice de um país, ou é dado pela taxa de câmbio da moeda local para USD, no momento t , caso se trate de um índice regional.

O factor 100 é o valor definido para o índice no momento inicial respeitante ao mercado.

⁶⁴ A informação constante neste Anexo foi obtida no site da Thompson Reuters[®], especificamente em http://thomsonreuters.com/content/financial/pdf/i_and_a/indices/country_region_index_methodology.pdf

ANEXO C

C.1. GRÁFICOS E ESTATÍSTICAS DESCRITIVAS DAS SÉRIES (LOG-NÍVEL E DL)

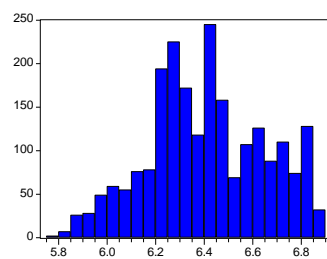
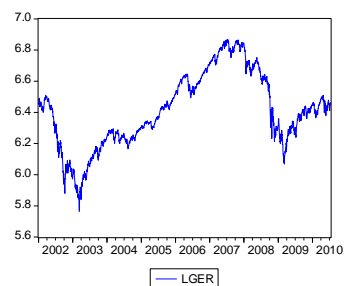
	LFR	LGER	LGR	LIT	LPT	LSP
Mean	7.470105	6.449528	7.197310	7.512126	5.179561	6.059481
Median	7.476767	6.422354	7.112466	7.477038	5.178012	6.048837
Maximum	7.864946	6.846061	7.895093	7.996287	5.676617	6.455057
Minimum	7.069968	6.068078	6.649917	6.970570	4.837551	5.650347
Std. Dev.	0.169828	0.169088	0.323631	0.212275	0.191606	0.173538
Skewness	0.047100	0.264134	0.459259	0.357579	0.471681	0.237189
Kurtosis	2.299069	2.246064	2.043598	2.427069	2.426862	2.263413
Jarque-Bera	13.77574	23.34118	48.42875	23.12676	33.55731	21.14083
Probability	0.001020	0.000009	0.000000	0.000010	0.000000	0.000026
Sum	4937.739	4263.138	4757.422	4965.515	3423.690	4005.317
Sum Sq. Dev.	19.03551	18.86994	69.12655	29.73998	24.23040	19.87627
Observations	661	661	661	661	661	661

TABELA 12 – Estatísticas descritivas das séries logaritmizadas (1:2008 – 7:2010)

	LFR	LGER	LGR	LIT	LPT	LSP
LFR	1.000000	0.981694	0.824418	0.971381	0.957911	0.954206
LGER	0.981694	1.000000	0.846511	0.974567	0.934460	0.939076
LGR	0.824418	0.846511	1.000000	0.922345	0.880638	0.917019
LIT	0.971381	0.974567	0.922345	1.000000	0.958437	0.979504
LPT	0.957911	0.934460	0.880638	0.958437	1.000000	0.960457
LSP	0.954206	0.939076	0.917019	0.979504	0.960457	1.000000

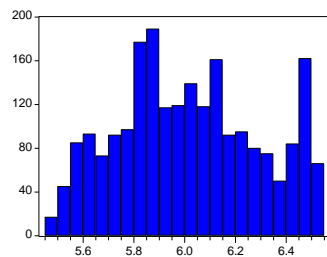
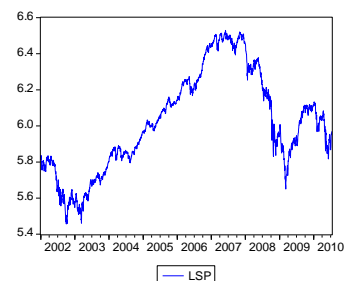
TABELA 13 – Matriz de Correlação entre as séries logaritmizadas (1:2008 – 7:2010)

ALEMANHA



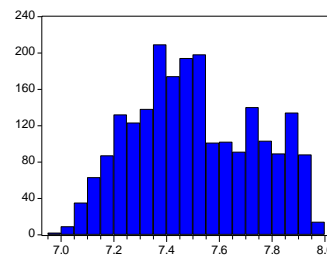
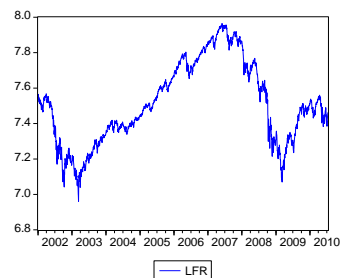
Series: LGER	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	6.410191
Median	6.407680
Maximum	6.868453
Minimum	5.765411
Std. Dev.	0.241559
Skewness	-0.005534
Kurtosis	2.361120
Jarque-Bera	37.86891
Probability	0.000000

ESPAÑA



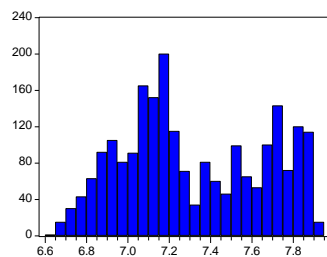
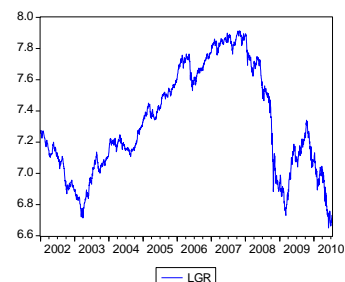
Series: LSP	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	6.019162
Median	6.002874
Maximum	6.529521
Minimum	5.456474
Std. Dev.	0.277769
Skewness	0.131382
Kurtosis	2.054841
Jarque-Bera	89.25980
Probability	0.000000

FRANÇA



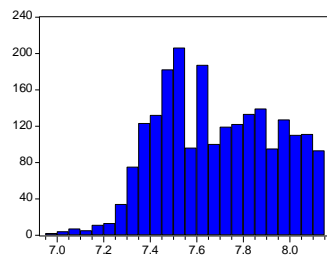
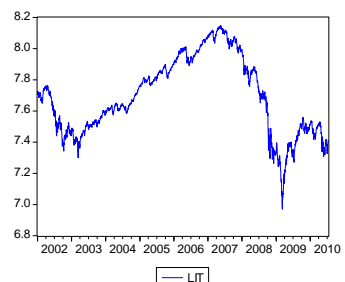
Series: LFR	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	7.509774
Median	7.490876
Maximum	7.964677
Minimum	6.961230
Std. Dev.	0.230538
Skewness	0.143291
Kurtosis	2.094143
Jarque-Bera	83.72597
Probability	0.000000

GRÉCIA



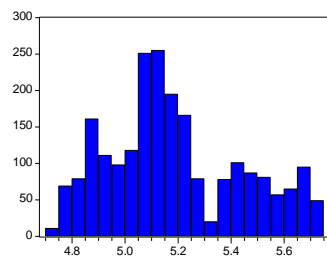
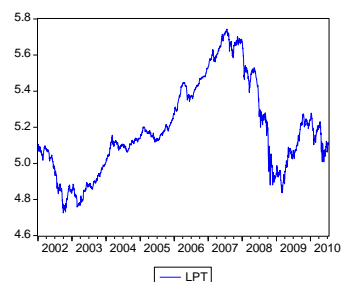
Series: LGR	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	7.318596
Median	7.224315
Maximum	7.914117
Minimum	6.649917
Std. Dev.	0.339633
Skewness	0.166900
Kurtosis	1.798601
Jarque-Bera	144.2061
Probability	0.000000

ITÁLIA



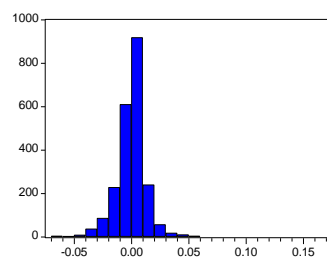
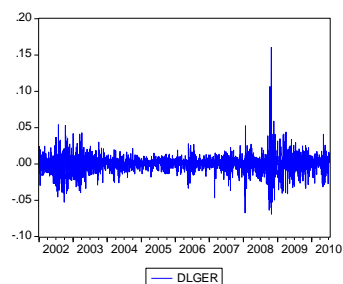
Series: LIT	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	7.690352
Median	7.671594
Maximum	8.149345
Minimum	6.970570
Std. Dev.	0.247236
Skewness	0.027129
Kurtosis	2.082230
Jarque-Bera	78.39646
Probability	0.000000

PORTUGAL

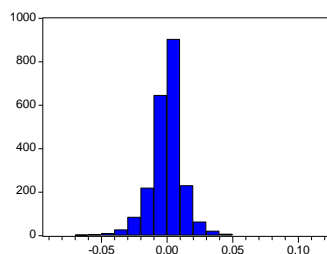
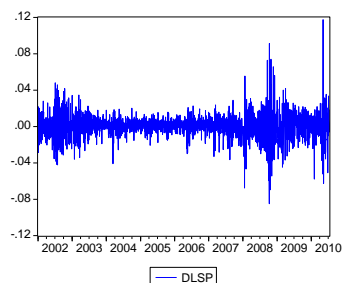


Series: LPT	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	5.188821
Median	5.139087
Maximum	5.743580
Minimum	4.725439
Std. Dev.	0.252986
Skewness	0.435492
Kurtosis	2.316125
Jarque-Bera	113.7391
Probability	0.000000

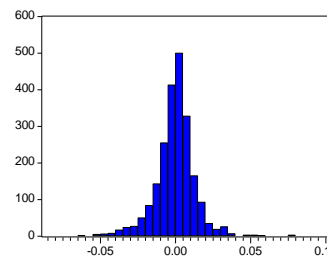
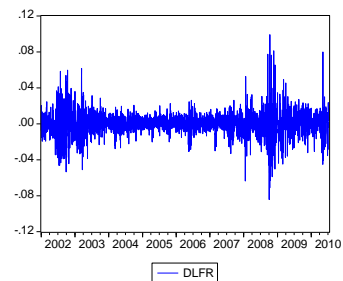
FIGURA 12 – Representação gráfica e estatísticas descritivas individualizada das séries logaritmizadas

ALEMANHA

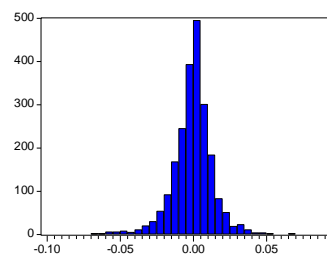
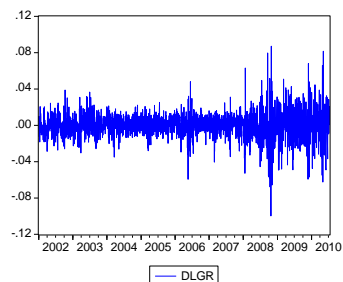
Series: DLGER	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	3.29e-06
Median	0.000574
Maximum	0.160461
Minimum	-0.069133
Std. Dev.	0.013201
Skewness	0.630878
Kurtosis	16.87462
Jarque-Bera	18002.50
Probability	0.000000

ESPAÑA

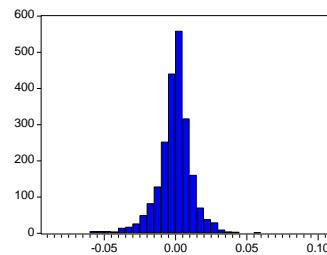
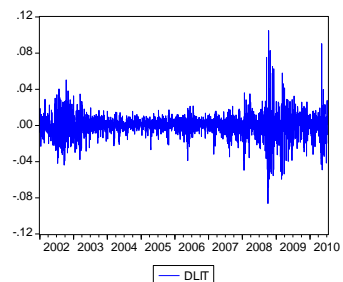
Series: DLSP	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	6.93e-05
Median	0.000707
Maximum	0.117492
Minimum	-0.084916
Std. Dev.	0.013157
Skewness	0.114900
Kurtosis	10.78516
Jarque-Bera	5626.355
Probability	0.000000

FRANÇA

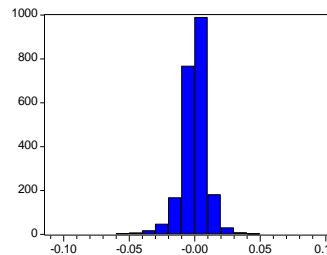
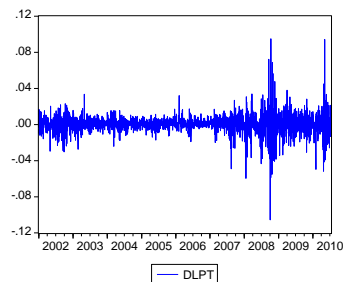
Series: DLFR	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	-3.89e-05
Median	0.000406
Maximum	0.099199
Minimum	-0.084287
Std. Dev.	0.013928
Skewness	0.071268
Kurtosis	8.801832
Jarque-Bera	3123.966
Probability	0.000000

GRÉCIA

Series: DLGR	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	-0.000235
Median	0.000000
Maximum	0.087081
Minimum	-0.099705
Std. Dev.	0.014284
Skewness	-0.270506
Kurtosis	8.178598
Jarque-Bera	2514.505
Probability	0.000000

ITÁLIA

Series: DLIT	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	-0.000135
Median	0.000379
Maximum	0.104822
Minimum	-0.086364
Std. Dev.	0.013084
Skewness	-0.007734
Kurtosis	10.63363
Jarque-Bera	5404.783
Probability	0.000000

PORTUGAL

Series: DLPT	
Sample 1/01/2002 7/13/2010	
Observations 2226	
Mean	6.19e-06
Median	0.000402
Maximum	0.095003
Minimum	-0.105610
Std. Dev.	0.010743
Skewness	-0.163640
Kurtosis	17.34787
Jarque-Bera	19103.57
Probability	0.000000

FIGURA 13 – Representação gráfica e estatísticas descritivas individualizada das séries

DL

C.2. CAUSALIDADE BIVARIADA DE GRANGER

REQUISITO:

O teste de causalidade de Granger assume que a informação relevante para a predição das respectivas variáveis X e Y está contida apenas nas séries de tempo sobre essas duas variáveis. Dessa forma, uma série de tempo estacionária X causa, no sentido de Granger, uma outra série estacionária Y se melhores predições estatisticamente significantes de Y podem ser obtidas ao incluirmos valores desfasados de X aos valores desfasados de Y . Em termos mais formais, o teste envolve estimar as seguintes regressões:

$$X_t = \sum_i a_i Y_{t-i} + \sum_i b_i X_{t-i} + u_{1t}$$

$$Y_t = \sum_i c_i Y_{t-i} + \sum_i d_i X_{t-i} + u_{2t}$$

onde u_{1t} e u_{2t} são os resíduos que assumimos serem não-correlacionados.

A primeira equação mostra que valores correntes de X estão relacionados a valores passados do próprio X assim como a valores desfasados de Y ; a segunda equação traduz um comportamento similar para a variável Y . Nada impede que as variáveis X e Y sejam representadas na forma de taxas de crescimento, o que aliás tem sido quase que a regra geral na literatura, uma vez que é difícil achar variáveis que sejam estacionárias em seus níveis.

Pairwise Granger Causality Tests

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Lags: 32

Null Hypothesis:	Obs	F-Statistic	Probability
DLGER does not Granger Cause DLFR	2194	6.20192	1.5E-24*
DLFR does not Granger Cause DLGER		2.69002	1.1E-06*
DLGR does not Granger Cause DLFR	2194	1.61710	0.01594*
DLFR does not Granger Cause DLGR		1.50670	0.03432*
DLIT does not Granger Cause DLFR	2194	1.57588	0.02137*
DLFR does not Granger Cause DLIT		1.37892	0.07717*
DLSP does not Granger Cause DLFR	2194	1.89944	0.00177*
DLFR does not Granger Cause DLSP		1.92990	0.00138*

DLPT does not Granger Cause DLFR	2194	1.40321	0.06660*
DLFR does not Granger Cause DLPT		1.73593	0.00657*
DLGR does not Granger Cause DLGER	2194	1.58005	0.02075*
DLGER does not Granger Cause DLGR		3.12503	1.1E-08*
DLIT does not Granger Cause DLGER	2194	3.41314	4.7E-10*
DLGER does not Granger Cause DLIT		4.75829	7.3E-17*
DLSP does not Granger Cause DLGER	2194	2.47373	9.5E-06*
DLGER does not Granger Cause DLSP		3.88389	2.2E-12*
DLPT does not Granger Cause DLGER	2194	1.68450	0.00972*
DLGER does not Granger Cause DLPT		3.01403	3.7E-08*
DLIT does not Granger Cause DLGR	2194	1.61217	0.01652*
DLGR does not Granger Cause DLIT		2.03529	0.00056*
DLSP does not Granger Cause DLGR	2194	1.62559	0.01500*
DLGR does not Granger Cause DLSP		2.07342	0.00040*
DLPT does not Granger Cause DLGR	2194	1.86699	0.00232*
DLGR does not Granger Cause DLPT		1.88647	0.00198*
DLSP does not Granger Cause DLIT	2194	2.44231	1.3E-05*
DLIT does not Granger Cause DLSP		2.01775	0.00065*
DLPT does not Granger Cause DLIT	2194	2.14978	0.00020*
DLIT does not Granger Cause DLPT		1.74343	0.00620*
DLPT does not Granger Cause DLSP	2194	2.35686	2.9E-05*
DLSP does not Granger Cause DLPT		1.68181	0.00991*

* Rejeita-se H_0 para um nível de significância de 5%

TABELA 14 – Estatísticas relativas ao teste de causalidade bivariada de Granger

ANEXO D

D.1. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2002 – 7:2010)

VAR Lag Order Selection Criteria

Endogenous variables: DLFR DLGER DLGR DLIT DLPT DLSP

Exogenous variables: C

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Included observations: 2200

Lag	LogL	LR	FPE	AIC	SC	HQ
0	45508.11	NA	4.37e-26	-41.36556	-41.35002	-41.35988
1	45663.39	309.5617	3.92e-26	-41.47399	-41.36524*	-41.43425
2	45755.07	182.2844	3.73e-26	-41.52461	-41.32265	-41.45081*
3	45808.63	106.1968	3.67e-26	-41.54057	-41.24541	-41.43272
4	45841.63	65.24072	3.68e-26	-41.53784	-41.14946	-41.39592
5	45901.56	118.1851	3.60e-26	-41.55960	-41.07801	-41.38363
6	45946.03	87.44438	3.57e-26	-41.56730	-40.99250	-41.35726
7	45998.54	102.9540	3.52e-26	-41.58231	-40.91430	-41.33821
8	46045.01	90.88488	3.48e-26	-41.59183	-40.83061	-41.31367
9	46071.90	52.43743	3.51e-26	-41.58355	-40.72912	-41.27133
10	46138.27	129.0420	3.42e-26	-41.61115	-40.66351	-41.26487
11	46174.49	70.23967	3.42e-26	-41.61135	-40.57050	-41.23101
12	46220.07	88.12863	3.39e-26	-41.62006	-40.48600	-41.20566
13	46269.84	95.97001	3.35e-26	-41.63258	-40.40530	-41.18412
14	46302.81	63.39582	3.35e-26	-41.62983	-40.30934	-41.14731
15	46337.90	67.28201	3.36e-26	-41.62900	-40.21531	-41.11242
16	46390.90	101.3316	3.31e-26	-41.64446	-40.13755	-41.09382
17	46435.93	85.83789	3.28e-26	-41.65267	-40.05255	-41.06797
18	46488.02	99.01938	3.23e-26	-41.66729	-39.97396	-41.04853
19	46515.32	51.73802	3.26e-26	-41.65938	-39.87284	-41.00656
20	46560.76	85.88915	3.23e-26*	-41.66797*	-39.78822	-40.98109
21	46584.82	45.33617	3.27e-26	-41.65711	-39.68415	-40.93617
22	46616.19	58.94157	3.28e-26	-41.65290	-39.58672	-40.89790
23	46640.69	45.91803	3.32e-26	-41.64245	-39.48307	-40.85339
24	46663.84	43.24940	3.35e-26	-41.63077	-39.37817	-40.80765
25	46691.39	51.31851*	3.38e-26	-41.62309	-39.27728	-40.76590

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

TABELA 15 – Estatísticas relativas à estimação do número de *lags* ótimo (1:2002 – 7:2010)

USANDO O SC – 1 LAG:

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Included observations: 2224

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 1

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	0	1	1	1	1
Max-Eig	0	1	1	1	1

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 16 – Estatísticas relativas ao sumário da ordem de cointegração para 1 lag (SC)
(1:2002 – 7:2010)

Date: ----- Time: -----

Sample (adjusted): 1/03/2002 7/13/2010

Included observations: 2224 after adjustments

Trend assumption: No deterministic trend (restricted constant)

Series: LFR LGER LGR LIT LPT LSP

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.020003	105.3202	103.8473	0.0398
At most 1	0.010628	60.38215	76.97277	0.4598
At most 2	0.007472	36.61788	54.07904	0.6434
At most 3	0.004709	19.93858	35.19275	0.7305
At most 4	0.002778	9.440366	20.26184	0.6944
At most 5	0.001462	3.254464	9.164546	0.5342

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.020003	44.93803	40.95680	0.0169
At most 1	0.010628	23.76427	34.80587	0.5397
At most 2	0.007472	16.67930	28.58808	0.6876
At most 3	0.004709	10.49821	22.29962	0.7944

At most 4	0.002778	6.185902	15.89210	0.7669		
At most 5	0.001462	3.254464	9.164546	0.5342		

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'S11*b=I):

LFR	LGER	LGR	LIT	LPT	LSP	C
-52.13758	27.07806	-6.304482	23.40203	1.804559	7.194401	31.43043
-7.796080	0.861882	-7.532456	12.13723	-12.22212	16.87513	-23.32215
-4.450872	-10.72406	-2.027671	1.183185	13.34948	1.907597	27.07520
1.559332	1.950485	-7.110069	8.340984	5.798392	-6.229312	-29.08643
3.419753	-4.637424	4.540468	0.365785	-6.464120	-0.603037	5.125522
1.113441	0.471702	-0.959016	-1.345844	-2.068000	1.725367	5.421074

Unrestricted Adjustment Coefficients (alpha):

D(LFR)	D(LGER)	D(LGR)	D(LIT)	D(LPT)	D(LSP)	D(C)
2.11E-05	-0.000167	0.000774	-0.000463	0.000400	3.71E-05	
-0.000702	7.72E-05	0.000917	-0.000260	0.000234	-2.73E-05	
0.000492	0.000357	0.000693	0.000283	0.000471	0.000121	
-0.000330	-0.000420	0.000624	-0.000253	0.000441	0.000116	
-0.000344	0.000259	0.000138	-0.000233	0.000482	3.95E-05	
-0.000107	-0.000323	0.000589	-0.000232	0.000507	-0.000102	

1 Cointegrating Equation(s):	Log likelihood	46180.14
------------------------------	----------------	----------

Normalized cointegrating coefficients (standard error in parentheses)

LFR	LGER	LGR	LIT	LPT	LSP	C
1.000000	-0.519358	0.120920	-0.448852	-0.034611	-0.137989	-0.602836
	(0.03987)	(0.03446)	(0.03644)	(0.05767)	(0.04817)	(0.14019)

Adjustment coefficients (standard error in parentheses)

D(LFR)	-0.001101
	(0.01514)
D(LGER)	0.036619
	(0.01451)
D(LGR)	-0.025661
	(0.01547)
D(LIT)	0.017191
	(0.01430)
D(LPT)	0.017948
	(0.01166)
D(LSP)	0.005588
	(0.01435)

TABELA 17 – Estatísticas relativas ao teste de cointegração de Johansen – 1 lag –
 (Intercept e no Trend⁶⁵) (1:2002 – 7:2010)

⁶⁵ Face à representação gráfica das séries, os dados apresentam restrições na constante (com *intercept*) e não evidenciam tendência determinística (no *trend*).

USANDO O AIC – 20 LAGS:

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Included observations: 2205

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 20

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Trace	0	0	0	0	0
Max-Eig	0	0	0	0	0

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 18 – Estatísticas relativas ao sumário da ordem de cointegração para 20 lags
(AIC) (1:2002 – 7:2010)

D.2. MODELOS VAR/VECM (1:2002 – 7:2010)

MODELO VECM: 6 variáveis, (séries do tipo $I(1)$), com 1 vector cointegrante considerando 1 lags (pelo SC)

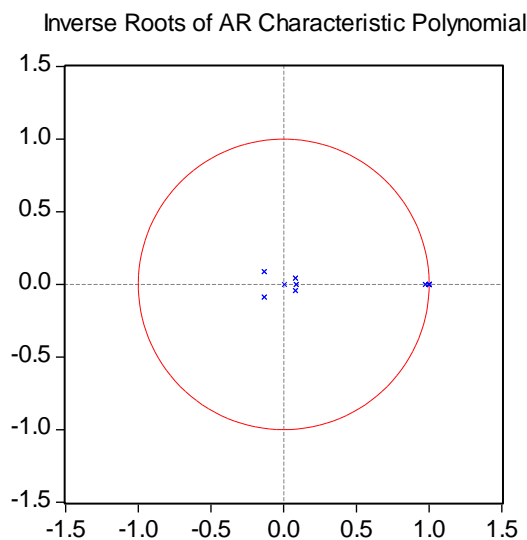


FIGURA 14 – Representação gráfica relativa à estabilidade do modelo VECM (1:2002 – 7:2010)

VEC Granger Causality/Block Exogeneity Wald Tests

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Included observations: 2224

Dependent variable: D(LFR)

Excluded	Chi-sq	df	Prob.
D(LGER)	78.86985	1	0.0000*
D(LGR)	2.399086	1	0.1214
D(LIT)	0.027959	1	0.8672
D(LPT)	0.245940	1	0.6199
D(LSP)	0.029092	1	0.8646
All	82.20316	5	0.0000

Dependent variable: D(LGER)

Excluded	Chi-sq	df	Prob.
D(LFR)	5.817852	1	0.0159
D(LGR)	10.54142	1	0.0012*
D(LIT)	0.450343	1	0.5022
D(LPT)	2.219708	1	0.1363
D(LSP)	0.237909	1	0.6257
All	28.55946	5	0.0000

Dependent variable: D(LGR)

Excluded	Chi-sq	df	Prob.
D(LFR)	17.38339	1	0.0000*
D(LGER)	75.19381	1	0.0000*
D(LIT)	0.034740	1	0.8521
D(LPT)	0.715900	1	0.3975
D(LSP)	0.787314	1	0.3749
All	81.87375	5	0.0000

Dependent variable: D(LIT)

Excluded	Chi-sq	df	Prob.
D(LFR)	14.57504	1	0.0001*
D(LGER)	52.03029	1	0.0000*
D(LGR)	5.690192	1	0.0171
D(LPT)	0.361199	1	0.5478
D(LSP)	8.66E-06	1	0.9977
All	58.79263	5	0.0000

Dependent variable: D(LPT)			
Excluded	Chi-sq	df	Prob.
D(LFR)	23.34960	1	0.0000*
D(LGER)	80.82221	1	0.0000*
D(LGR)	0.620665	1	0.4308
D(LIT)	1.627194	1	0.2021
D(LSP)	0.479853	1	0.4885
All	85.69677	5	0.0000

Dependent variable: D(LSP)			
Excluded	Chi-sq	df	Prob.
D(LFR)	29.01213	1	0.0000*
D(LGER)	56.82627	1	0.0000*
D(LGR)	2.119076	1	0.1455
D(LIT)	0.013340	1	0.9080
D(LPT)	1.125915	1	0.2886
All	67.89462	5	0.0000

* Rejeita-se H_0 para um nível de significância de 1%

TABELA 19 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VECM (1:2002 – 7:2010)

MODELO VAR: 6 variáveis (séries do tipo $I(1)$), não cointegradas considerando 20 lags (pelo AIC)

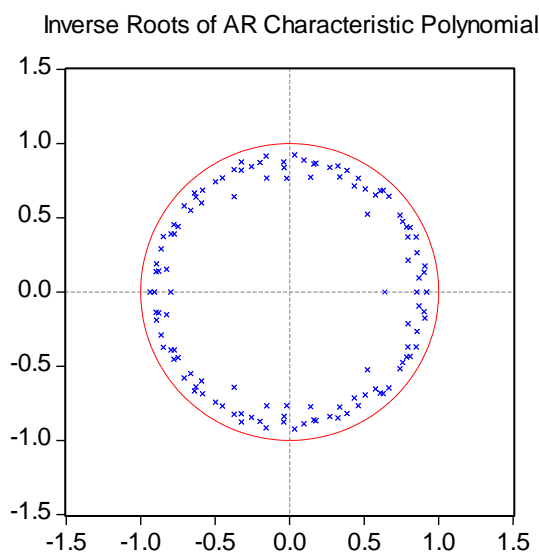


FIGURA 15 – Representação gráfica relativa à estabilidade do modelo VAR (1:2002 – 7:2010)

VAR Granger Causality/Block Exogeneity Wald Tests

Date: ----- Time: -----

Sample: 1/01/2002 7/13/2010

Included observations: 2205

Dependent variable: DLFR

Excluded	Chi-sq	df	Prob.
DLGER	148.2264	20	0.0000*
DLGR	36.69497	20	0.0127
DLIT	37.94856	20	0.0090*
DLPT	41.75000	20	0.0030*
DLSP	53.75618	20	0.0001*
All	336.8579	100	0.0000

Dependent variable: DLGER

Excluded	Chi-sq	df	Prob.
DLFR	40.38593	20	0.0045*
DLGR	47.45103	20	0.0005*
DLIT	51.39179	20	0.0001*
DLPT	38.44997	20	0.0078*
DLSP	22.28654	20	0.3252
All	239.3841	100	0.0000

Dependent variable: DLGR

Excluded	Chi-sq	df	Prob.
DLFR	52.63441	20	0.0001*
DLGER	125.6747	20	0.0000*
DLIT	35.01624	20	0.0200*
DLPT	20.06975	20	0.4536
DLSP	29.23765	20	0.0832
All	253.9693	100	0.0000

Dependent variable: DLIT

Excluded	Chi-sq	df	Prob.
DLFR	41.51623	20	0.0032*
DLGER	128.5976	20	0.0000*
DLGR	37.86232	20	0.0092*
DLPT	55.09680	20	0.0000*
DLSP	50.87378	20	0.0002*
All	329.5080	100	0.0000

Dependent variable: DLPT

Excluded	Chi-sq	df	Prob.
DLFR	64.16998	20	0.0000*
DLGER	158.3321	20	0.0000*
DLGR	42.01779	20	0.0028*
DLIT	33.67930	20	0.0284*
DLSP	46.52895	20	0.0007*
All	356.9845	100	0.0000

Dependent variable: DLSP

Excluded	Chi-sq	df	Prob.
DLFR	61.71593	20	0.0000*
DLGER	138.3895	20	0.0000*
DLGR	38.89512	20	0.0069*
DLIT	32.17617	20	0.0415
DLPT	48.77791	20	0.0003*
All	336.9620	100	0.0000

* Rejeita-se H_0 para um nível de significância de 1%

TABELA 20 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VAR
(1:2002 – 7:2010)

D.3. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2002 – 1:2008)

VAR Lag Order Selection Criteria

Endogenous variables: LFR LGER LGR LIT LPT LSP

Exogenous variables: C

Date: ----- Time: -----

Sample: 1/01/2002 1/01/2008

Included observations: 1534

Lag	LogL	LR	FPE	AIC	SC	HQ
0	13672.83	NA	7.36e-16	-17.81856	-17.79769	-17.81079
1	34138.54	40744.64	1.99e-27	-44.45443	-44.30834*	-44.40006
2	34232.95	187.2176	1.84e-27	-44.53058	-44.25928	-44.42962*
3	34273.67	80.42476	1.83e-27*	-44.53673*	-44.14021	-44.38918
4	34305.85	63.31297	1.84e-27	-44.53175	-44.01001	-44.33760
5	34335.18	57.47311	1.86e-27	-44.52305	-43.87610	-44.28231
6	34369.26	66.51326	1.86e-27	-44.52055	-43.74838	-44.23321
7	34396.80	53.53502	1.88e-27	-44.50952	-43.61213	-44.17558
8	34427.33	59.10313	1.90e-27	-44.50238	-43.47977	-44.12185
9	34455.81	54.92082	1.92e-27	-44.49258	-43.34476	-44.06545
10	34478.96	44.46505	1.95e-27	-44.47583	-43.20279	-44.00211
11	34521.01	80.42405	1.93e-27	-44.48371	-43.08546	-43.96340

12	34563.00	79.98566	1.92e-27	-44.49153	-42.96805	-43.92461
13	34590.90	52.93206	1.94e-27	-44.48097	-42.83228	-43.86746
14	34625.96	66.23245	1.94e-27	-44.47974	-42.70583	-43.81964
15	34654.95	54.54056	1.96e-27	-44.47060	-42.57148	-43.76390
16	34683.30	53.10453	1.98e-27	-44.46062	-42.43628	-43.70733
17	34701.61	34.15924	2.03e-27	-44.43756	-42.28800	-43.63767
18	34731.99	56.44352	2.04e-27	-44.43023	-42.15545	-43.58374
19	34758.11	48.33918	2.07e-27	-44.41736	-42.01737	-43.52428
20	34785.14	49.78483	2.09e-27	-44.40566	-41.88045	-43.46598
21	34812.46	50.12478	2.12e-27	-44.39435	-41.74392	-43.40807
22	34838.17	46.96141	2.15e-27	-44.38093	-41.60529	-43.34806
23	34870.49	58.78688	2.16e-27	-44.37613	-41.47527	-43.29667
24	34895.05	44.46863	2.19e-27	-44.36121	-41.33514	-43.23515
25	34924.14	52.45057	2.21e-27	-44.35220	-41.20091	-43.17955
26	34944.76	37.02646	2.26e-27	-44.33216	-41.05564	-43.11291
27	34970.27	45.59239	2.29e-27	-44.31847	-40.91675	-43.05263
28	35001.95	56.37538	2.31e-27	-44.31284	-40.78589	-43.00040
29	35034.30	57.32051*	2.32e-27	-44.30808	-40.65592	-42.94904
30	35058.74	43.11627	2.35e-27	-44.29301	-40.51563	-42.88738
31	35083.23	43.00458	2.39e-27	-44.27800	-40.37541	-42.82577
32	35098.13	26.06359	2.46e-27	-44.25050	-40.22269	-42.75168

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

TABELA 21 – Estatísticas relativas à estimação do número de *lags* óptimo (1:2002 – 1:2008)

USANDO O SC – 1 LAG:

Date: ----- Time: -----

Sample: 1/01/2002 1/01/2008

Included observations: 1564

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 1

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	2	1	1	2	2
Max-Eig	2	1	1	2	2

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 22 – Estatísticas relativas ao sumário da ordem de cointegração para 1 *lag* (SC)
(1:2002 – 1:2008)

USANDO O AIC – 3 LAGS:

Date: ----- Time: -----

Sample: 1/01/2002 1/01/2008

Included observations: 1562

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 3

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Trace	2	1	1	1	1
Max-Eig	1	0	0	1	0

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 23 – Estatísticas relativas ao sumário da ordem de cointegração para 3 lags (AIC)
(1:2002 – 1:2008)

D.4. ANÁLISE DA ORDEM DE COINTEGRAÇÃO (1:2008 – 7:2010)

VAR Lag Order Selection Criteria

Endogenous variables: LFR LGER LGR LIT LPT LSP

Exogenous variables: C

Date: ----- Time: -----

Sample: 1/01/2008 7/13/2010

Included observations: 629

Lag	LogL	LR	FPE	AIC	SC	HQ
0	5760.94115882982	NA	4.55e-16	-18.29870	-18.25631	-18.28223
1	12002.416702663	12344.03	1.23e-24	-38.02994	-37.73320*	-37.91467*
2	12067.954175297	128.3659	1.12e-24	-38.12386	-37.57276	-37.90978
3	12129.5996095603	119.5667	1.03e-24	-38.20540	-37.39995	-37.89252
4	12163.0464166246	64.23489	1.04e-24	-38.19729	-37.13748	-37.78560
5	12199.1405279117	68.63046	1.04e-24	-38.19759	-36.88342	-37.68709
6	12234.5354835507	66.62580	1.04e-24	-38.19566	-36.62714	-37.58636
7	12271.750654453	69.34210	1.04e-24	-38.19953	-36.37665	-37.49142
8	12316.0975604692	81.78444	1.01e-24*	-38.22607*	-36.14884	-37.41916
9	12349.5455029152	61.04648	1.02e-24	-38.21795	-35.88637	-37.31224
10	12371.1988033832	39.10676	1.07e-24	-38.17233	-35.58640	-37.16782
11	12405.3553742899	61.03654	1.08e-24	-38.16647	-35.32618	-37.06315
12	12433.0140944183	48.89745	1.11e-24	-38.13995	-35.04531	-36.93782
13	12468.0078152488	61.19729	1.11e-24	-38.13675	-34.78775	-36.83582
14	12500.5261959369	56.24801	1.13e-24	-38.12568	-34.52233	-36.72594
15	12536.5390362595	61.60543	1.13e-24	-38.12572	-34.26802	-36.62718
16	12567.3133266451	52.05699	1.15e-24	-38.10910	-33.99705	-36.51176
17	12601.3651188643	56.95149	1.16e-24	-38.10291	-33.73650	-36.40676
18	12631.7896162861	50.30442	1.19e-24	-38.08518	-33.46442	-36.29023
19	12680.1989746759	79.11736	1.14e-24	-38.12464	-33.24952	-36.23088

20	12700.9137909929	33.45986	1.21e-24	-38.07604	-32.94656	-36.08347
21	12741.6934679953	65.09189	1.19e-24	-38.09124	-32.70741	-35.99987
22	12759.5357811368	28.13923	1.27e-24	-38.03350	-32.39532	-35.84333
23	12781.0446919432	33.51150	1.34e-24	-37.98742	-32.09489	-35.69845
24	12802.9049020894	33.64179	1.41e-24	-37.94246	-31.79557	-35.55468
25	12849.5916355583	70.95790	1.37e-24	-37.97644	-31.57520	-35.48986
26	12867.4603228019	26.81723	1.46e-24	-37.91879	-31.26319	-35.33340
27	12889.7296704102	32.99687	1.54e-24	-37.87513	-30.96518	-35.19094
28	12933.0867952204	63.41583	1.52e-24	-37.89853	-30.73422	-35.11552
29	12975.1590034577	60.73381*	1.50e-24	-37.91783	-30.49917	-35.03603
30	12996.8210718959	30.85725	1.59e-24	-37.87225	-30.19923	-34.89163
31	13022.5886791635	36.21393	1.66e-24	-37.83971	-29.91234	-34.76029
32	13046.8721439474	33.66484	1.74e-24	-37.80246	-29.62073	-34.62423

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

TABELA 24 – Estatísticas relativas à estimação do número de *lags* óptimo (1:2008 – 7:2010)

USANDO O SC – 1 LAG:

Date: ----- Time: -----

Sample: 1/01/2008 7/13/2010

Included observations: 659

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 1

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Trace	0	0	0	0	0
Max-Eig	1	1	1	0	0

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 25 – Estatísticas relativas ao sumário da ordem de cointegração para 1 *lag* (SC)
(1:2008 – 7:2010)

Date: ----- Time: -----

Sample (adjusted): 1/03/2008 7/13/2010

Included observations: 659 after adjustments

Trend assumption: No deterministic trend (restricted constant)

Series: LFR LGER LGR LIT LPT LSP

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.062932	100.7439	103.8473	0.0788
At most 1	0.031324	57.90946	76.97277	0.5628
At most 2	0.023500	36.93678	54.07904	0.6272
At most 3	0.014919	21.26572	35.19275	0.6447
At most 4	0.012836	11.35985	20.26184	0.5078
At most 5	0.004310	2.846222	9.164546	0.6098

Trace test indicates no cointegration at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.062932	42.83448	40.95680	0.0304
At most 1	0.031324	20.97268	34.80587	0.7491
At most 2	0.023500	15.67106	28.58808	0.7679
At most 3	0.014919	9.905872	22.29962	0.8424
At most 4	0.012836	8.513627	15.89210	0.4877
At most 5	0.004310	2.846222	9.164546	0.6098

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

LFR	LGER	LGR	LIT	LPT	LSP	C
-7.052214	34.60164	6.003294	-45.75264	6.391075	12.22513	22.69928
57.13672	-19.02228	10.78011	-31.27844	-15.26963	-1.224177	-60.16817
9.317637	-14.09370	2.178668	8.116925	11.91540	-24.86375	33.30299
14.44436	2.841320	4.968755	0.926983	-1.233571	-21.91425	-29.52350
1.240291	6.480780	3.305833	3.088688	-13.61123	-4.857134	1.588722
-6.307715	1.333936	1.499961	-0.605307	1.781215	-1.158668	30.91637

Unrestricted Adjustment Coefficients (alpha):

D(LFR)	0.002752	-0.000601	0.001214	-0.000901	0.000107	0.000522
D(LGER)	0.001128	0.000285	0.002035	-0.000879	-0.000396	0.000373
D(LGR)	0.001922	-0.001670	0.001781	-0.001140	0.000617	-0.000362
D(LIT)	0.003748	-0.000118	0.001312	-0.000659	7.05E-05	0.000293
D(LPT)	0.002061	0.000311	0.001104	-0.000894	0.000930	0.000292
D(LSP)	0.002758	-0.000600	0.001677	-0.000233	0.000500	0.000473

1 Cointegrating Equation(s):		Log likelihood		12627.88		
Normalized cointegrating coefficients (standard error in parentheses)						
LFR	LGER	LGR	LIT	LPT	LSP	C
1.000000	-4.906493	-0.851264	6.487700	-0.906251	-1.733517	-3.218746
	(0.77868)	(0.22057)	(1.10047)	(0.43227)	(0.72413)	(1.17357)
Adjustment coefficients (standard error in parentheses)						
D(LFR)	-0.019406					
	(0.00483)					
D(LGER)	-0.007953					
	(0.00474)					
D(LGR)	-0.013554					
	(0.00570)					
D(LIT)	-0.026433					
	(0.00503)					
D(LPT)	-0.014538					
	(0.00444)					
D(LSP)	-0.019448					
	(0.00494)					

TABELA 26 – Estatísticas relativas ao teste de cointegração de Johansen – 1 lag –
(*Intercept* e no *Trend*⁶⁶) (1:2008 – 7:2010)

USANDO O AIC – 8 LAGS:

Date: ----- Time: -----

Sample: 1/01/2008 7/13/2010

Included observations: 652

Series: LFR LGER LGR LIT LPT LSP

Lags interval: 1 to 8

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	0	0	0	0	0
Max-Eig	0	0	0	0	0

*Critical values based on MacKinnon-Haug-Michelis (1999)

TABELA 27 – Estatísticas relativas ao sumário da ordem de cointegração para 8 lags (AIC)
(1:2008 – 7:2010)

⁶⁶ Face à representação gráfica das séries, os dados apresentam restrições na constante (com *intercept*) e não evidenciam tendência determinística (*no trend*).

D.5. MODELOS VAR/VECM (1:2008 – 7:2010)

MODELO VECM: 6 variáveis, (séries do tipo $I(1)$), com 1 vector cointegrante considerando 1 lags (pelo SC)

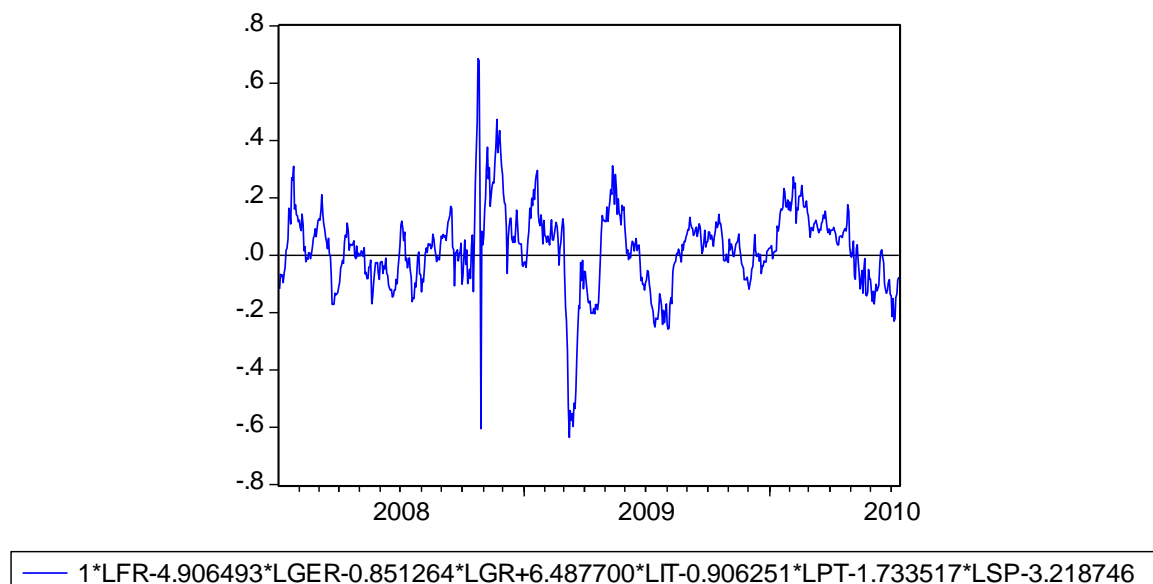


FIGURA 16 – Representação gráfica da relação de longo prazo (1:2008 – 7:2010)

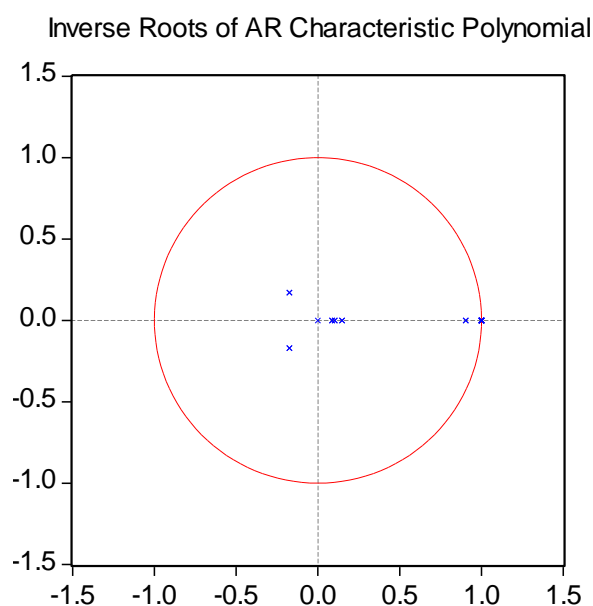


FIGURA 17 – Representação gráfica relativa à estabilidade do modelo VECM (1:2008 – 7:2010)

VEC Granger Causality/Block Exogeneity Wald Tests

Date: ----- Time: -----

Sample: 1/01/2008 7/13/2010

Included observations: 659

Dependent variable: D(LFR)

Excluded	Chi-sq	df	Prob.
D(LGER)	22.57993	1	0.0000*
D(LGR)	0.727532	1	0.3937
D(LIT)	0.221081	1	0.6382
D(LPT)	0.370168	1	0.5429
D(LSP)	0.214178	1	0.6435
All	24.23516	5	0.0002

Dependent variable: D(LGER)

Excluded	Chi-sq	df	Prob.
D(LFR)	6.657998	1	0.0099*
D(LGR)	6.567404	1	0.0104
D(LIT)	0.844766	1	0.3580
D(LPT)	0.560565	1	0.4540
D(LSP)	0.206725	1	0.6493
All	15.89953	5	0.0071

Dependent variable: D(LGR)

Excluded	Chi-sq	df	Prob.
D(LFR)	2.505820	1	0.1134
D(LGER)	21.71767	1	0.0000*
D(LIT)	0.608648	1	0.4353
D(LPT)	0.542170	1	0.4615
D(LSP)	1.303956	1	0.2535
All	26.23637	5	0.0001

Dependent variable: D(LIT)

Excluded	Chi-sq	df	Prob.
D(LFR)	5.356736	1	0.0206
D(LGER)	13.83129	1	0.0002*
D(LGR)	1.088404	1	0.2968
D(LPT)	0.603424	1	0.4373
D(LSP)	0.294694	1	0.5872
All	16.63011	5	0.0053

Dependent variable: D(LPT)

Excluded	Chi-sq	df	Prob.
D(LFR)	4.480878	1	0.0343
D(LGER)	22.09427	1	0.0000*
D(LGR)	0.699626	1	0.4029
D(LIT)	0.365567	1	0.5454
D(LSP)	0.020080	1	0.8873
All	24.21830	5	0.0002

Dependent variable: D(LSP)

Excluded	Chi-sq	df	Prob.
D(LFR)	11.38979	1	0.0007*
D(LGER)	16.24441	1	0.0001*
D(LGR)	1.179239	1	0.2775
D(LIT)	0.365031	1	0.5457
D(LPT)	0.859857	1	0.3538
All	23.84503	5	0.0002

* Rejeita-se H_0 para um nível de significância de 1%

TABELA 28 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VECM (1:2008 – 7:2010)

MODELO VAR: 6 variáveis (séries do tipo $I(1)$), não cointegradas considerando 8 lags (pelo AIC)

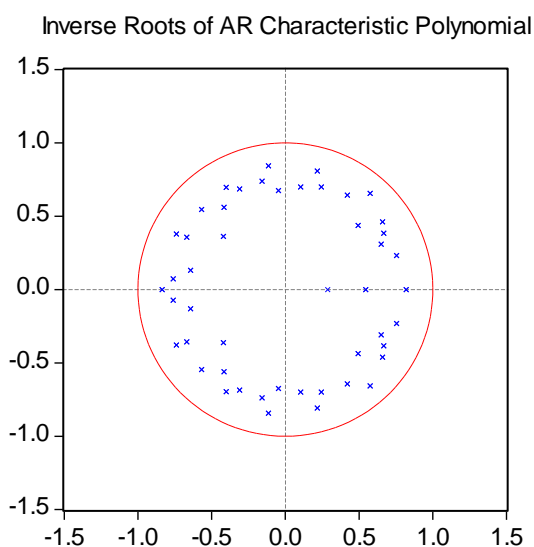


FIGURA 18 – Representação gráfica relativa à estabilidade do modelo VAR (1:2008 – 7:2010)

VAR Granger Causality/Block Exogeneity Wald Tests

Date: ----- Time: -----

Sample: 1/01/2008 7/13/2010

Included observations: 652

Dependent variable: DLFR

Excluded	Chi-sq	df	Prob.
DLGER	76.30589	8	0.0000*
DLGR	12.52432	8	0.1293
DLIT	9.214376	8	0.3245
DLPT	19.35504	8	0.0131
DLSP	9.125907	8	0.3318
All	131.9283	40	0.0000

Dependent variable: DLGER

Excluded	Chi-sq	df	Prob.
DLFR	26.23538	8	0.0010*
DLGR	24.43747	8	0.0019*
DLIT	17.57315	8	0.0247
DLPT	19.63066	8	0.0118
DLSP	7.600659	8	0.4734
All	112.5379	40	0.0000

Dependent variable: DLGR

Excluded	Chi-sq	df	Prob.
DLFR	24.03610	8	0.0023*
DLGER	62.76922	8	0.0000*
DLIT	9.115366	8	0.3327
DLPT	14.93626	8	0.0604
DLSP	12.04662	8	0.1491
All	100.8002	40	0.0000

Dependent variable: DLIT

Excluded	Chi-sq	df	Prob.
DLFR	29.87891	8	0.0002*
DLGER	73.97019	8	0.0000*
DLGR	12.72689	8	0.1216
DLPT	23.68516	8	0.0026*
DLSP	10.96105	8	0.2039

All	134.5397	40	0.0000
Dependent variable: DLPT			
Excluded	Chi-sq	df	Prob.
DLFR	24.46511	8	0.0019*
DLGER	64.28952	8	0.0000*
DLGR	16.94443	8	0.0307
DLIT	6.134567	8	0.6322
DLSP	13.06335	8	0.1097
All	134.7917	40	0.0000
Dependent variable: DLSP			
Excluded	Chi-sq	df	Prob.
DLFR	37.13522	8	0.0000*
DLGER	68.20396	8	0.0000*
DLGR	10.99312	8	0.2021
DLIT	7.846175	8	0.4486
DLPT	17.44492	8	0.0258
All	131.6598	40	0.0000

* Rejeita-se H_0 para um nível de significância de 1%

TABELA 29 – Estatísticas relativas ao estudo da causalidade de Granger no modelo VAR
(1:2008 – 7:2010)

ANEXO E

E.1. ANÁLISE BAYESIANA DA ORDEM DE COINTEGRAÇÃO (1:2002 – 7:2010)

USANDO O SC – 2 LAG:

Structural VAR --- version 0.45				
Copyright © Anders Warne, 2001-2011				

Endogenous variables: pt sp it fr ger gr				
Estimation Sample: 3:1 - 2226:1				
Initial Sample: 1:1 - 2:1				
Number of lags: 2 for endogenous variables				
Number of observations: 2224				
Deterministic variables: const				

BAYESIAN ANALYSIS				
=====				
Posterior Cointegration Rank Probabilities				
Rank	P(Rank Data)	log[P(Data Rank)]	P(Rank)	
0	1.0000	39892.1257	0.1429	
1	0.0000	39880.0881	0.1429	
2	0.0000	39874.0174	0.1429	
3	0.0000	39844.5633	0.1429	
4	0.0000	39849.6536	0.1429	
5	0.0000	39832.3252	0.1429	
6	0.0000	39832.5508	0.1429	
NOTE: The calculations are based on the Fractional Marginal Likelihood (FML) for a diffuse prior on the parameters conditional on the rank and equal prior probabilities for each rank. The fraction used is minimal, 0.05, and is the relative part of the sample used to "train" the improper (diffuse) prior into a proper distribution.				
Posterior Cointegration Rank Probabilities				
Rank	P(Rank Data)	log[P(Data Rank)]	P(Rank)	st.err. log[P(Data Rank)]
0	0.9999	30880.8583	0.1429	NaN
1	0.0001	30871.4357	0.1429	0.3657
2	0.0000	30864.8714	0.1429	0.1926
3	0.0000	30855.7280	0.1429	0.2240
4	0.0000	30846.6839	0.1429	0.8652
5	0.0000	30839.3140	0.1429	0.0785
6	0.0000	30830.9463	0.1429	NaN

NOTE: The calculations are based on the **Marginal Likelihood (ML)**, given by $P(\text{Data}|\text{Rank})$. It has been computed from the basic marginal likelihood identity with all parameters evaluated at the posterior mode. Numerical standard errors of the log marginal likelihood are based on the Newey-West correction, taking potential autocorrelation of the beta draws into account. Requires that the partial Gibbs sampler has converged.

Standard errors for the full rank and zero rank models do not exist as these are exact.

Hyperparameters for Prior Distribution

~~~~~

Baseline shrinkage: 0.3000  
 Lag length shrinkage: 1.0000  
 Off-diagonal/correlation shrinkage: 0.0000  
 Deterministic shrinkage: 1.0000  
 Alpha shrinkage: 1.0000  
 Omega prior type: A matrix shrinkage times identity  
 A matrix shrinkage: 1.0000  
 q parameter: 8

\*\*\*\*\*

#### Cointegration rank 0

=====

#### Summary statistics from the Gibbs sampler

~~~~~

Total draws: 1100
 Burn-in draws: 100
 Considered draws: 1000
 Total discarded draws: 0
 Discarded explosive draws: 0
 Draw saving frequency: 100.0%
 Saved draws: 1000

Posterior Mode Properties

~~~~~

Max absolute eigenvalue of companion matrix: 0.0040  
 Value of log prior density: -5045.9974  
 Value of log data density: 36328.3321  
 Value of log posterior density: 401.4764

\*\*\*\*\*

**TABELA 30** – Estimação bayesiana da ordem de cointegração 2 lag (SC) (1:2002 – 7:2010)

USANDO O AIC – 19 LAGS:

```

*****
                        Structural VAR --- version 0.45
                        Copyright © Anders Warne, 2001-2011
*****

Endogenous variables:  pt sp it fr ger gr
Estimation Sample:    20:1 - 2226:1
Initial Sample:       1:1 - 19:1
Number of lags:       19 for endogenous variables
Number of observations: 2207
Deterministic variables: const
*****

                        BAYESIAN ANALYSIS
                        =====

Posterior Cointegration Rank Probabilities

Rank  P(Rank|Data)  log[P(Data|Rank)]  P(Rank)
-----
0      1.0000      37399.6206         0.1429
1      0.0000      37380.2461         0.1429
2      0.0000      37375.1217         0.1429
3      0.0000      37343.9797         0.1429
4      0.0000      37348.2985         0.1429
5      0.0000      37330.3392         0.1429
6      0.0000      37330.8759         0.1429

NOTE: The calculations are based on the Fractional Marginal Likelihood (FML) for a diffuse prior on the
parameters conditional on the rank and equal prior probabilities for each rank. The fraction used is
minimal, 0.05, and is the relative part of the sample used to "train" the improper (diffuse) prior into a
proper distribution.

*****

Posterior Cointegration Rank Probabilities

Rank  P(Rank|Data)  log[P(Data|Rank)]  P(Rank)  st.err. log[P(Data|Rank)]
-----
0      0.9998      30601.4562         0.1429      NaN
1      0.0002      30593.0746         0.1429      0.1050
2      0.0000      30585.3651         0.1429      0.1875
3      0.0000      30576.8128         0.1429      0.2437
4      0.0000      30568.6583         0.1429      0.1782
5      0.0000      30559.9999         0.1429      0.0882
6      0.0000      30551.8048         0.1429      NaN

NOTE: The calculations are based on the Marginal Likelihood (ML), given by P(Data|Rank). It has been
computed from the basic marginal likelihood identity with all parameters evaluated at the posterior mode.
Numerical standard errors of the log marginal likelihood are based on the Newey-West correction, taking
potential autocorrelation of the beta draws into account. Requires that the partial Gibbs sampler has
converged.
Standard errors for the full rank and zero rank models do not exist as these are exact.

*****

Hyperparameters for Prior Distribution
~~~~~

Baseline shrinkage: 0.3000
Lag length shrinkage: 1.0000

```

```

Off-diagonal/correlation shrinkage: 0.0000
Deterministic shrinkage: 1.0000
Alpha shrinkage: 1.0000
Omega prior type: A matrix shrinkage times identity
A matrix shrinkage: 1.0000
q parameter: 8

```

```

```

```

Cointegration rank 0

```

```

=====

```

```

Summary statistics from the Gibbs sampler

```

```

~~~~~

```

```

Total draws: 1100
Burn-in draws: 100
Considered draws: 1000
Total discarded draws: 0
Discarded explosive draws: 0
Draw saving frequency: 100.0%
Saved draws: 1000

```

```

Posterior Mode Properties

```

```

~~~~~

```

```

Max absolute eigenvalue of companion matrix: 0.5717
Value of log prior density: -1476.0893
Value of log data density: 36253.2357
Value of log posterior density: 4175.6902

```

```

```

**TABELA 31** – Estimação bayesiana da ordem de cointegração 19 lags (AIC) (1:2002 – 7:2010)

## E.2. ANÁLISE BAYESIANA DA ORDEM DE COINTEGRAÇÃO (1:2008 – 7:2010)

USANDO O SC – 1 LAG:

```

Structural VAR --- version 0.45
Copyright © Anders Warne, 2001-2011

Endogenous variables: pt sp it fr ger gr
Estimation Sample: 9:1 - 661:1
Initial Sample: 8:1 - 8:1
Number of lags: 1 for endogenous variables
Number of observations: 660
Deterministic variables: const

```



## BAYESIAN ANALYSIS

## Posterior Cointegration Rank Probabilities

| Rank | P(Rank Data) | log[P(Data Rank)] | P(Rank) |
|------|--------------|-------------------|---------|
| 0    | 0.2286       | 9290.5869         | 0.1429  |
| 1    | 0.6968       | 9291.7016         | 0.1429  |
| 2    | 0.0746       | 9289.4672         | 0.1429  |
| 3    | 0.0000       | 9266.7090         | 0.1429  |
| 4    | 0.0000       | 9271.1868         | 0.1429  |
| 5    | 0.0000       | 9257.9492         | 0.1429  |
| 6    | 0.0000       | 9258.0529         | 0.1429  |

NOTE: The calculations are based on the **Fractional Marginal Likelihood (FML)** for a diffuse prior on the parameters conditional on the rank and equal prior probabilities for each rank. The fraction used is minimal, 0.02, and is the relative part of the sample used to "train" the improper (diffuse) prior into a proper distribution.

## Posterior Cointegration Rank Probabilities

| Rank | P(Rank Data) | log[P(Data Rank)] | P(Rank) | st.err. log[P(Data Rank)] |
|------|--------------|-------------------|---------|---------------------------|
| 0    | 0.9933       | 6868.1027         | 0.1429  | NaN                       |
| 1    | 0.0066       | 6863.0955         | 0.1429  | 0.0717                    |
| 2    | 0.0000       | 6857.8785         | 0.1429  | 0.2452                    |
| 3    | 0.0000       | 6854.5400         | 0.1429  | 0.4170                    |
| 4    | 0.0000       | 6848.1363         | 0.1429  | 0.5481                    |
| 5    | 0.0000       | 6843.8036         | 0.1429  | 0.2386                    |
| 6    | 0.0000       | 6837.3024         | 0.1429  | NaN                       |

NOTE: The calculations are based on the **Marginal Likelihood (ML)**, given by  $P(\text{Data}|\text{Rank})$ . It has been computed from the basic marginal likelihood identity with all parameters evaluated at the posterior mode. Numerical standard errors of the log marginal likelihood are based on the Newey-West correction, taking potential autocorrelation of the beta draws into account. Requires that the partial Gibbs sampler has converged.

Standard errors for the full rank and zero rank models do not exist as these are exact.

## Hyperparameters for Prior Distribution

```

Hyperparameters for Prior Distribution
~~~~~
Baseline shrinkage:          0.3000
Lag length shrinkage:       1.0000
Off-diagonal/correlation shrinkage: 0.0000
Deterministic shrinkage:    1.0000
Alpha shrinkage:            1.0000
Omega prior type:           A matrix shrinkage times identity
A matrix shrinkage:         1.0000
q parameter:                 8
*****

```

## Cointegration rank 0

## Summary statistics from the Gibbs sampler

```

~~~~~
Total draws: 1100
Burn-in draws: 100
Considered draws: 1000
Total discarded draws: 0

```

Discarded explosive draws: 0  
 Draw saving frequency: 100.0%  
 Saved draws: 1000

#### Posterior Mode Properties

~~~~~  
 Max absolute eigenvalue of companion matrix: 1.0000  
 Value of log prior density: -1477.9102  
 Value of log data density: 8558.7192  
 Value of log posterior density: 212.7062

\*\*\*\*\*

#### Cointegration rank 1

=====

#### Summary statistics from the Gibbs sampler

~~~~~  
 Total draws: 1855  
 Burn-in draws: 100  
 Considered draws: 1755  
 Total discarded draws: 755  
 Discarded explosive draws: 755  
 Draw saving frequency: 100.0%  
 Saved draws: 1000

#### Posterior Mean Cointegration Space Estimator

~~~~~  
 1 -0.1469 0.0788 0.8703 -0.3038 -0.2834 -0.2054

with projective Frobenius span variation: 0.8352

#### Normalized beta based on the Identifying Restrictions:

1 -0.1688 0.0905 1.0000 -0.3490 -0.3257 -0.2360

#### Partial Posterior Mean Cointegration Space Estimator

~~~~~  
 1 -0.0521 -0.0361 0.8550 -0.2777 -0.4024 -0.1611

with projective Frobenius span variation: 0.8853

#### Normalized beta based on the Identifying Restrictions:

1 -0.0610 -0.0422 1.0000 -0.3249 -0.4707 -0.1885

#### Posterior Mode Cointegration Space Estimator

~~~~~  
 1 -3.6808 4.1952 14.0333 -2.1724 -10.2431 -2.3512

#### Normalized beta based on the Identifying Restrictions:

1 -0.2623 0.2989 1.0000 -0.1548 -0.7299 -0.1675

#### Posterior Mode Properties

~~~~~  
 Max absolute eigenvalue of companion matrix: 0.9878  
 Value of log prior density: -1472.8200  
 Value of log data density: 8564.4668  
 Value of log posterior density: 228.5513

\*\*\*\*\*

**TABELA 32** – Estimação bayesiana da ordem de cointegração 1 lag (SC) (1:2008 – 7:2010)

## USANDO O AIC – 8 LAGS:

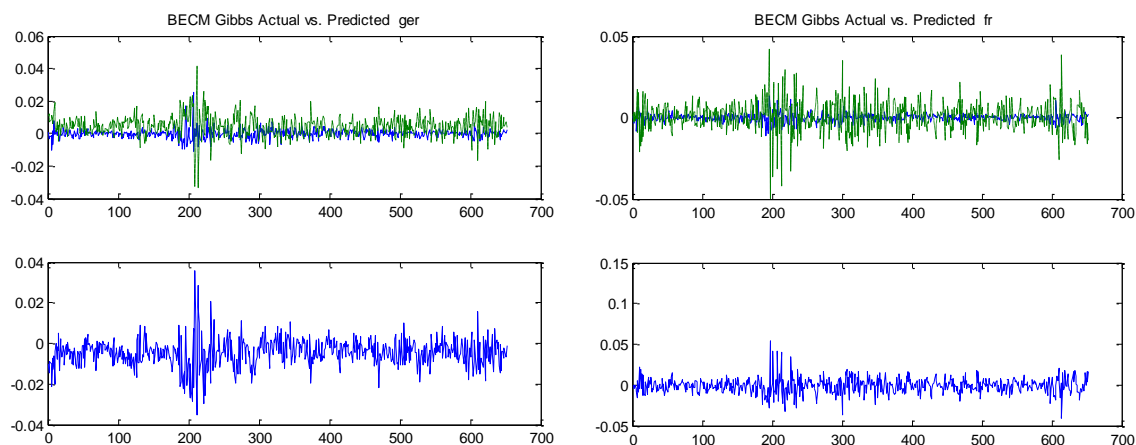
|                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                   |                   |         |                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|-------------------|---------|---------------------------|
| *****                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| Structural VAR --- version 0.45                                                                                                                                                                                                                                                                                                                                                                                                         |                                   |                   |         |                           |
| Copyright © Anders Warne, 2001-2011                                                                                                                                                                                                                                                                                                                                                                                                     |                                   |                   |         |                           |
| *****                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| Endogenous variables: pt sp it fr ger gr                                                                                                                                                                                                                                                                                                                                                                                                |                                   |                   |         |                           |
| Estimation Sample: 9:1 - 661:1                                                                                                                                                                                                                                                                                                                                                                                                          |                                   |                   |         |                           |
| Initial Sample: 1:1 - 8:1                                                                                                                                                                                                                                                                                                                                                                                                               |                                   |                   |         |                           |
| Number of lags: 8 for endogenous variables                                                                                                                                                                                                                                                                                                                                                                                              |                                   |                   |         |                           |
| Number of observations: 653                                                                                                                                                                                                                                                                                                                                                                                                             |                                   |                   |         |                           |
| Deterministic variables: const                                                                                                                                                                                                                                                                                                                                                                                                          |                                   |                   |         |                           |
| *****                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| BAYESIAN ANALYSIS                                                                                                                                                                                                                                                                                                                                                                                                                       |                                   |                   |         |                           |
| =====                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| Posterior Cointegration Rank Probabilities                                                                                                                                                                                                                                                                                                                                                                                              |                                   |                   |         |                           |
| Rank                                                                                                                                                                                                                                                                                                                                                                                                                                    | P(Rank Data)                      | log[P(Data Rank)] | P(Rank) |                           |
| 0                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.9999                            | 8566.2876         | 0.1429  |                           |
| 1                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0001                            | 8556.4333         | 0.1429  |                           |
| 2                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 8550.9088         | 0.1429  |                           |
| 3                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 8526.5147         | 0.1429  |                           |
| 4                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 8531.2518         | 0.1429  |                           |
| 5                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 8516.9271         | 0.1429  |                           |
| 6                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 8517.0094         | 0.1429  |                           |
| NOTE: The calculations are based on the Fractional Marginal Likelihood (FML) for a diffuse prior on the parameters conditional on the rank and equal prior probabilities for each rank. The fraction used is minimal, 0.08, and is the relative part of the sample used to "train" the improper (diffuse) prior into a proper distribution.                                                                                             |                                   |                   |         |                           |
| Posterior Cointegration Rank Probabilities                                                                                                                                                                                                                                                                                                                                                                                              |                                   |                   |         |                           |
| Rank                                                                                                                                                                                                                                                                                                                                                                                                                                    | P(Rank Data)                      | log[P(Data Rank)] | P(Rank) | st.err. log[P(Data Rank)] |
| 0                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.9944                            | 6867.8022         | 0.1429  | NaN                       |
| 1                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0056                            | 6862.6150         | 0.1429  | 0.0913                    |
| 2                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 6857.3481         | 0.1429  | 0.3493                    |
| 3                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 6854.7006         | 0.1429  | 0.3758                    |
| 4                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 6847.5425         | 0.1429  | 0.5258                    |
| 5                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 6843.8432         | 0.1429  | 0.2285                    |
| 6                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.0000                            | 6836.9991         | 0.1429  | NaN                       |
| NOTE: The calculations are based on the Marginal Likelihood (ML), given by P(Data Rank). It has been computed from the basic marginal likelihood identity with all parameters evaluated at the posterior mode. Numerical standard errors of the log marginal likelihood are based on the Newey-West correction, taking potential autocorrelation of the beta draws into account. Requires that the partial Gibbs sampler has converged. |                                   |                   |         |                           |
| Standard errors for the full rank and zero rank models do not exist as these are exact.                                                                                                                                                                                                                                                                                                                                                 |                                   |                   |         |                           |
| *****                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| Hyperparameters for Prior Distribution                                                                                                                                                                                                                                                                                                                                                                                                  |                                   |                   |         |                           |
| ~~~~~                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                   |                   |         |                           |
| Baseline shrinkage:                                                                                                                                                                                                                                                                                                                                                                                                                     | 0.3000                            |                   |         |                           |
| Lag length shrinkage:                                                                                                                                                                                                                                                                                                                                                                                                                   | 1.0000                            |                   |         |                           |
| Off-diagonal/correlation shrinkage:                                                                                                                                                                                                                                                                                                                                                                                                     | 0.0000                            |                   |         |                           |
| Deterministic shrinkage:                                                                                                                                                                                                                                                                                                                                                                                                                | 1.0000                            |                   |         |                           |
| Alpha shrinkage:                                                                                                                                                                                                                                                                                                                                                                                                                        | 1.0000                            |                   |         |                           |
| Omega prior type:                                                                                                                                                                                                                                                                                                                                                                                                                       | A matrix shrinkage times identity |                   |         |                           |

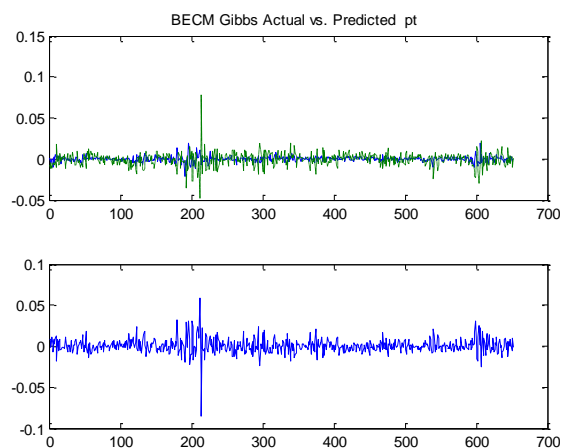
|                                              |           |
|----------------------------------------------|-----------|
| A matrix shrinkage:                          | 1.0000    |
| q parameter:                                 | 8         |
| *****                                        |           |
| Cointegration rank 0                         |           |
| =====                                        |           |
| Summary statistics from the Gibbs sampler    |           |
| ~~~~~                                        |           |
| Total draws:                                 | 600       |
| Burn-in draws:                               | 100       |
| Considered draws:                            | 500       |
| Total discarded draws:                       | 0         |
| Discarded explosive draws:                   | 0         |
| Draw saving frequency:                       | 100.0%    |
| Saved draws:                                 | 500       |
| Posterior Mode Properties                    |           |
| ~~~~~                                        |           |
| Max absolute eigenvalue of companion matrix: | 0.2770    |
| Value of log prior density:                  | -399.9787 |
| Value of log data density:                   | 8662.3940 |
| Value of log posterior density:              | 1394.6131 |
| *****                                        |           |

**TABELA 33** – Estimação bayesiana da ordem de cointegração 8 lags (AIC) (1:2008 – 7:2010)

### E.3. MODELOS BVAR/VECM (1:2008 – 7:2010)

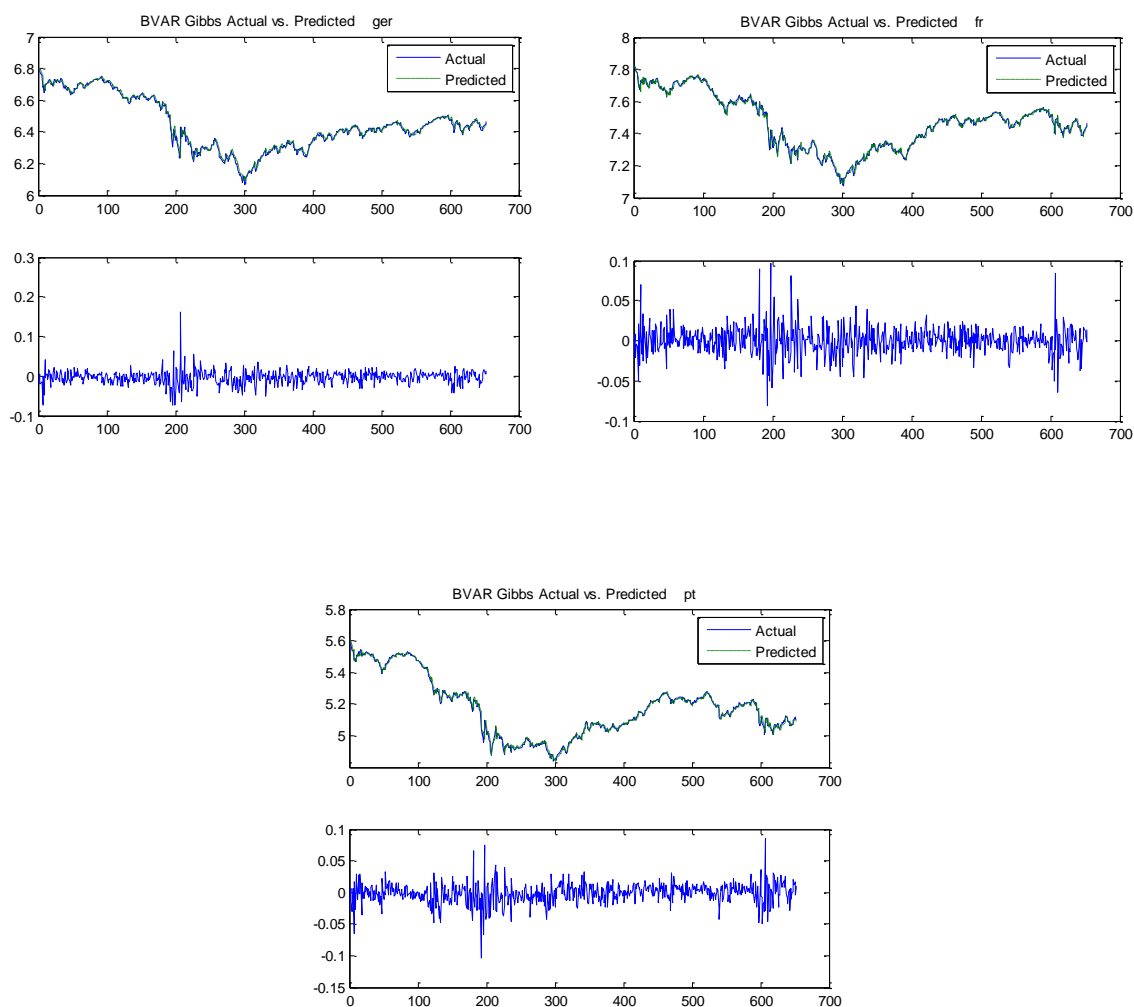
MODELO BVECM: (considerando a existência de um vector cointegrante)





**FIGURA 19** – Representações gráficas relativas ao modelo BVECM

MODELO BVAR: (considerando a inexistência de vectores cointegrantes)



**FIGURA 20** – Representações gráficas relativas ao modelo BVAR